

Interaction in Multidimensional Contingency Tables: An Information Theoretic Approach^{1,2}

H. H. Ku

Institute for Basic Standards, National Bureau of Standards,
Washington, D.C. 20234

and

S. Kullback

The George Washington University,
Washington, D.C. 20006

(May 10, 1968)

The problem of interaction in multidimensional contingency tables is investigated from the viewpoint of information theory as developed by Kullback. The hypothesis of no r th-order interaction is defined in the sense of an hypothesis of "generalized" independence of classifications with fixed r th order marginal restraints. For a three-way table, with given cell probabilities π_{ijk} , the minimum discrimination information for a contingency table with marginals $p_{ij\cdot}$, $p_{\cdot jk}$, and $p_{i\cdot k}$ is given by the set of cell probabilities $p_{ijk}^* = a_{ij}b_{jk}c_{ik}\pi_{ijk}$ where a_{ij} , b_{jk} , and c_{ik} are functions of the given marginal probabilities, that is, $\ln(p_{ijk}^*/\pi_{ijk}) = \ln a_{ij} + \ln b_{jk} + \ln c_{ik}$, representing no second-order interaction. The minimum discrimination information statistic, asymptotically distributed as χ^2 with appropriate degrees of freedom is

$$2\sum_{ijk} x_{ijk} \ln x_{ijk} - 2\sum_{ijk} x_{ijk} \ln x_{ijk}^* \geq 0$$

where x_{ijk} are the observed cell frequencies and x_{ijk}^* are the "no interaction" cell frequencies uniquely determined by a simple convergent iteration process of the marginals on π_{ijk} . For lower order marginal restraints the usual independence hypotheses are generated when π_{ijk} are taken to be the cell probabilities under uniform distribution. It is shown that the set p_{ijk}^* satisfies definitions of no second order interaction in a $2 \times 2 \times 2$ table given by Bartlett and no interaction in a $r \times s \times t$ table by Roy and Kastenbaum, and is also related to that given by Good. Results of application to the analysis of some "classical" three-dimensional contingency tables are given, together with full details for two four-dimensional examples.

Key Words: Contingency tables; estimation of cell frequencies from marginals; generalized independence; hypothesis testing; information theory; interaction; second-order interaction.

Introduction

In the last decade, a series of related papers appeared in various publications on the analysis of multiway contingency tables. The topic of particular interest was that of the definition and treatment of higher-order interaction. Among these papers, we cite, for instance, Roy and Kastenbaum [1956],³ Plackett [1962], Darroch [1962], Birch [1963], Good [1963], and Goodman [1964a,

¹ Presented by title to the Institute of Mathematical Statistics, Columbus, Ohio 23–25 March 1967 under the title Interaction in Multi-dimensional Contingency Tables; Abstract in Annals of Mathematical Statistics, Vol. 38, 297 (1967).

² Supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under Grant AF-AFOSR 932-65.

³ Figures in brackets indicate the literature references at the end of this paper.

1964b]. Also related were The Estimation of Probabilities by Good [1965], two reports by Bhapkar and Koch [1965, 1966], and their recent publication in Technometrics [1968].

In these papers, various aspects of the problem were treated and solutions offered. Consequently, results given are usually concerned with individual parts of the problem, depending on the author's motivation and interest. We propose to show, that some of these results can be unified through the use of information theory. The formulation, comparisons, and proofs will be given in section 2, and applications to the analysis of multidimensional contingency tables in section 3, following a historical review in section 1. First, however, we shall give a brief description of the problem and the notations that will be used.

Consider a random sample of n independent observations, where each observation can be classified by m criteria of classifications, say: row (R), column (C), and depth (D), for instance. Suppose there are r categories in the row classification, c categories in the column classification, and d categories in the depth classification, then there are exactly $r \times c \times d$ cells in the three dimensional table. Each observation by definition must fall in one of the cells, say, the ijk th cell, with probability p_{ijk} , where $i = 1, 2, \dots, r, j = 1, 2, \dots, c, k = 1, 2, \dots, d$, are the subscripts relating the cell to the categories in the row, column, and depth classifications respectively.

Let x_{ijk} represent the observed frequency of the ijk th cell in the sample, so that $\sum_{ijk} x_{ijk} = n$. Then we have summarized the data in the form of a three-way contingency table with cell probabilities p_{ijk} , $\sum_{ijk} p_{ijk} = 1$.

Summing over categories over one classification or two classifications, we obtain two-way and one-way marginal tables respectively, or in symbols:

$$x_{ij.} = \sum_k x_{ijk}, \quad x_{i..} = \sum_j x_{ij.} = \sum_{jk} x_{ijk},$$

and the like. Corresponding to these marginal frequency tables, we have similar marginal tables for the probabilities.

In the analysis of contingency table we are usually interested in the relationship between one classification and one or more of the other classifications. Suppose the row classification represents the response of an experiment on animals, the column classification types of treatment, and the depth classification a distinguishable characteristic of the sampled individuals, sex, for instance. Then in many respects the hypotheses of interest are analogous to those of independence and correlation in normal multivariate analysis, e.g.,

1. Response is independent of treatment, or

$$H_0: p_{ij.} = p_{i..} p_{.j.}$$

This case corresponds to simple correlation. That is, H_0 corresponds to the hypothesis that response and treatment are uncorrelated.

2. Response is independent of treatment and sex, or

$$H_0: p_{ijk} = p_{i..} p_{.jk.}$$

This case corresponds to multiple correlation.

3. Response is independent of treatment, given the sex, or

$$H_0: p_{ijk} = \frac{p_{i.k} p_{.jk.}}{p_{..k}}$$

This case corresponds to partial correlation.

Of course not all contingency tables can be interpreted in such a straightforward manner. In some cases all three classifications can be considered as responses; then we may be interested in the independence or associations among these responses. In other cases a classification may

be viewed either as a factor or a response. For convenience, we may group all the concepts of association, dependence, etc., under the general term of interaction. No interaction between treatment and sex appears to be a more acceptable phrase than independence between treatment and sex since independence is usually reserved to express the relationship between two random variables. We may also say that the interaction between response and treatment does not interact with sex, meaning the degree of association between response and treatment is the same for both sexes.

Here we come to grips with a concept which gives rise to the idea of second-order interaction (sometimes called a three-factor interaction when applied to a three-way table). The definition, method of analysis, and the interpretation of the second-order and higher-order interactions have been the source of controversy for a number of years. It is probably worthwhile to review the historical development of the problem and to summarize the various schools of thought on the subject to gain a proper perspective for the present treatment. A brief account will be given in the next section.

1. Historical Background.

1.1. Formulation of the No-Interaction Hypothesis

The first use of the "no second-order interaction" hypothesis as relating to a $2 \times 2 \times 2$ contingency table was due to Bartlett [1935]. The concept remained dormant for a number of years. Lewis [1962], in his excellent review on the subject, lamented that "... there is still no coordinated information available, and the treatment of these tables is still widely neglected in standard text books."

Bartlett's definition was mainly intuitive. Since then, there have been several attempts to arrive at a logical, consistent, and intuitively acceptable definition that could be derived from within a wider framework of hypothesis formulation. The main lines of thought can be grouped into the following classifications:

1. Bartlett's original definition and its extensions.
2. Simpson, Plackett, and Darroch's formulations based on symmetrical functions of the cell probabilities.
3. Good's definition based on maximum entropy and Goodman's modification.

1.2. Bartlett's Definition and Its Extension by Roy and Kastenbaum

Bartlett defined his term, formulated the hypothesis, proposed the statistic and suggested a method for the solution in less than 25 lines! To use the author's words, "The testing of independence in a 2×2 table with fixed marginal totals, may be regarded as testing the significance of the interaction between the two classifications Corresponding to the hypothesis to be tested in an ordinary fourfold table (i.e., a 2×2 table) of observed numbers n_1, n_2, n_3, n_4 that $p_1p_4 = p_2p_3$, we require to test the hypothesis (of no second-order interaction) that

$$(1.1) \quad p_1p_4p_6p_7 = p_2p_3p_5p_8."$$

$$(i.e., p_{111}p_{122}p_{212}p_{221} = p_{112}p_{121}p_{211}p_{222}.)$$

Thus for a 2×2 table with fixed marginal totals, Bartlett's definition of no first-order interaction implies and is implied by independence of the two classifications. Furthermore, he assumed that the cross-product ratio type of hypothesis can be extended to define second-order interaction for $2 \times 2 \times 2$ tables. It is remarkable that his definition remains the preferred one to this date and the same hypothesis has been arrived at by others through different approaches.

Bartlett's definition, however, becomes complicated when the categories within a classification are more than two—a difficulty acknowledged by him in the latter part of his paper. Proper inter-

pretation of results of the test also becomes difficult. Moreover, the solution requires lengthy iterative computation for the solution of $(r-1)(c-1)(d-1)$ simultaneous third degree equations, where r , c , and d are respectively, the number of categories in the row (R), column (C), and depth (D) classifications.

Norton [1945] extended Bartlett's definition to $2 \times 2 \times d$ tables, and devised an iterative procedure for solving these systems of equations. Roy and Kastenbaum [1956], commenting that "Bartlett's and Norton's papers do not give any indication of the mechanism behind the formula for the hypothesis of no interaction . . .", derived a set of "no interaction constraints" in an $r \times c \times d$ table in the form of

$$(1.2) \quad \frac{p_{rcd}p_{ijd}}{p_{ica}p_{rjd}} = \frac{p_{rck}p_{ijk}}{p_{ick}p_{rjk}} \quad \begin{array}{l} i=1, 2, \dots, (r-1) \\ \text{for } j=1, 2, \dots, (c-1) \\ k=1, 2, \dots, (d-1). \end{array}$$

The set of constraints reduces to (1.1) for a $2 \times 2 \times 2$ table.

The "mechanism" used by Roy and Kastenbaum is based on the fact that the two hypotheses

$$H_1: p_{i \cdot k} = p_{i \cdot} p_{\cdot \cdot k}$$

$$H_2: p_{ij \cdot} = p_{i \cdot} p_{\cdot j \cdot}$$

will not usually imply

$$H: p_{ijk} = p_{i \cdot} p_{\cdot jk}$$

in a three-way contingency table. The "no interaction" hypothesis is required to generate the set of constraints such that these constraints, when superimposed on $H_1 \cap H_2$ should imply H . The result is the set of constraints in (1.2). In contrast to Bartlett, Roy and Kastenbaum called (1.2) the hypothesis of "no interaction" or "no first-order interaction." The extension of this concept to the hypothesis of "no second-order interaction" in a four-way table was only indicated in their paper.

1.3. Simpson, Plackett, and Darroch's Formulation

Simpson [1951] required the definition of "no second-order interaction" to be symmetrical with respect to the three attributes of a $2 \times 2 \times 2$ table. If some function $\psi(p_{111}, p_{121}, p_{211}, p_{221})$ is chosen to measure the association of classifications R and C in D , then the function must be such that the equation

$$\psi(p_{111}, p_{121}, p_{211}, p_{221}) = \psi(p_{112}, p_{122}, p_{212}, p_{222})$$

implies and is implied by the equations

$$\psi(p_{111}, p_{211}, p_{112}, p_{212}) = \psi(p_{121}, p_{221}, p_{122}, p_{222})$$

and

$$\psi(p_{111}, p_{121}, p_{112}, p_{122}) = \psi(p_{211}, p_{221}, p_{212}, p_{222}).$$

He showed that the function $\psi = \frac{p_{121}p_{211}}{p_{111}p_{221}}$ or the cross-product ratio used by Bartlett, satisfies this requirement. Hence, Bartlett's definition for a $2 \times 2 \times 2$ table was accepted. The uniqueness of this function was not discussed.

In a footnote to Simpson's paper, the editor suggested that "This paper should be read in conjunction with the following paper by H. O. Lancaster." Lancaster [1951] defined the second-

order interaction by the partition of the chi-square statistic X^2 ; i.e., it is defined as the difference between the total X^2 for testing complete independence of the three classifications, and the sum of the three components corresponding to tests for independence in each of the three marginal tables.

Plackett [1962] compared Simpson's definition with Lancaster's definition [1951] and showed that the latter does not always satisfy the condition of symmetry. He accepted Roy and Kastenbaum's definition given in (1.2) for an $r \times c \times d$ table, and extended the analysis of log-frequencies [Woolf 1955] to such tables as an alternative method of analysis which is computationally easier than the solution of $(r-1)(c-1)(d-1)$ simultaneous equations of the third degree.

Darroch [1962] made an explicit comparison of the definitions of interaction in multiway contingency tables and in the analysis of variance. He found that there are resemblances between the two definitions but "that interactions in contingency tables enjoyed only a few of the fortuitously simple properties of interactions in the analysis of variance." The main point he made (also made by Roy and Kastenbaum) was that a natural symmetrical definition of "no second-order interaction"

$$(1.3) \quad p_{ijk} = \frac{p_{\cdot jk} p_{i \cdot k} p_{ij \cdot}}{p_{i \cdot \cdot} p_{\cdot j \cdot} p_{\cdot \cdot k}}$$

necessarily imposes constraints on the marginal probabilities $p_{ij \cdot}$, $p_{\cdot jk}$, $p_{i \cdot k}$, i.e.,

$$\sum_k p_{ijk} = p_{ij \cdot} = \frac{p_{ij \cdot}}{p_{i \cdot \cdot} p_{\cdot j \cdot}} \sum_k \frac{p_{i \cdot k} p_{\cdot jk}}{p_{\cdot \cdot k}},$$

or

$$\sum_k \frac{p_{i \cdot k} p_{\cdot jk}}{p_{\cdot \cdot k}} = p_{i \cdot \cdot} p_{\cdot j \cdot}$$

for all i, j , and the like. This is of course undesirable since the condition for "no second-order interaction" should relate p_{ijk} to any given set of marginal probabilities and should not place restrictions on the latter [cf. p. 172 Kullback 1959].

Consequently Darroch defined a "perfect three-way table" as one for which condition (1.3), and the resulting restraints on the marginal probabilities, are satisfied exactly. He concluded further that "in imperfect tables it is not possible to express p_{ijk} in terms of simple functions of $p_{ij \cdot}$, $p_{i \cdot k}$, and $p_{\cdot jk}$ when there is no second-order interaction." The existence and uniqueness of the set p_{ijk} as the solution of (1.2) for any given set of mutually consistent marginal probabilities was conjectured for $r \times c \times d$ tables and proved for the $2 \times 2 \times 2$ case. The search for a simple formulation in terms of parameters which are implicitly defined by the marginal probabilities led Darroch to define

$$p_{ijk} = \mu \alpha_{jk} \beta_{ki} \gamma_{ij}$$

where

$$\sum_k \alpha_{jk} = \sum_i \beta_{ki} = \sum_j \gamma_{ij} = 1,$$

and

$$\mu \sum_{ijk} \alpha_{jk} \beta_{ki} \gamma_{ij} = 1;$$

and showed that

$$\mu = 1, \alpha_{jk} = \frac{p_{\cdot jk}}{p_{\cdot j \cdot}}, \beta_{ki} = \frac{p_{i \cdot k}}{p_{i \cdot \cdot}}, \text{ and } \gamma_{ij} = \frac{p_{ij \cdot}}{p_{i \cdot \cdot}}.$$

Since there is no solution in closed form to the maximum likelihood equations for the parameters under hypothesis of no second-order interaction, unless the observed table happens to be perfect, Darroch suggested an iterative solution and gave a numerical illustration using the example given by Kastenbaum and Lamphiear [1959].

It is of interest to note that Darroch suggested the likelihood ratio test based on

$$Z_{RCD} = 2 \sum_{ijk} n_{ijk} \ln \left(\frac{n_{ijk}}{n \hat{\mu} \hat{\alpha}_{jk} \hat{\beta}_{ki} \hat{\gamma}_{ij}} \right)$$

which is asymptotically distributed as χ^2 with $(r-1) \times (c-1) \times (d-1)$ degrees of freedom.

Birch [1963] accepted (1.2) as the definition of no second-order interaction in a 3-way table and discussed maximum likelihood estimation of expected frequencies for many-way tables under different hypotheses. He also proved the conjecture by Darroch that the expected frequencies in a three-way table are uniquely given by the marginal totals if the expected frequencies are known to be positive and to satisfy the hypothesis of no second-order interaction. Thus, given any set of positive integers n_{ijk} he showed that there is one and only one set of positive numbers n_{ijk}^* that satisfies the equations

$$n_{ij.}^* = n_{ij.}, \quad n_{i.k}^* = n_{i.k}, \quad \text{and} \quad n_{.jk}^* = n_{.jk},$$

and also the conditions given by the no second-order interaction hypothesis expressed by (1.2).

1.4. Good's Formulation

The formulations of the hypothesis of no second-order interaction summarized up to now are basically extensions of Bartlett's. Justifications for such formulations are given in a number of ways: (1) residually as the difference between the independence of one classification (R) with the other two classifications (CD) and the two independence hypotheses ($R \times C$) and ($R \times D$); (2) by symmetry requirements; (3) by analogy with analysis of variance. Lancaster [1951] also proposed a formulation based strictly on the analogy of the partition of χ^2 to the analysis of variance. The shortcomings of his method were discussed by Lewis [1962] and Plackett [1961] and will not be repeated here.

Good [1963] proposed to use the principle of maximum entropy as a heuristic principle for the generation of null hypotheses, with main application to m -dimensional contingency tables. Three versions of this principle are given in his paper. We quote here his Principle of Minimal Discriminability: "Let X be a random variable whose distribution is subject to some set of restraints. Suppose that, before the restraints were known, there was some distribution that seemed reasonable to entertain as a null hypothesis, called an initially ausgezeichnet hypothesis. This hypothesis is perhaps refuted by the constraints. Then, in view of the restraints, entertain the null hypothesis that, if true, can be discriminated from the ausgezeichnet hypothesis at the minimum rate, i.e., for which the expected weight of evidence per observation is least."

Numerous examples and theorems are given in Good's paper. By using his principle, it is shown that for an m -dimensional $2 \times 2 \times \dots \times 2$ contingency table $(p_i) = (p_{i_1 i_2 \dots i_m})$, $i_1, i_2, \dots, i_m = 0, 1$, and with all the marginal probabilities down to $(m-1)$ -way assigned, the null hypothesis to be tested is

$$\prod_i^{|i| \text{ even}} p_i = \prod_i^{|i| \text{ odd}} p_i$$

where $|i| = i_1 + i_2 + \dots + i_m$. The expression reduces to (1.1) when $m=3$.

Good also generalized the definition to that of no r th-order and all higher-order interactions in an m -dimensional contingency table with a complete set of r th-order restraints by means of discrete Fourier transforms of the logarithms of probabilities. However, the interactions so defined are usually complex valued unless the categories within each classification are equal to two. Goodman

[1964] followed the definition by Good but proposed a test that yields real valued interactions. Goodman's proposed test is based on Wald's criterion [1943] and unrestricted maximum likelihood estimates, and is essentially an extension of the tests proposed by Plackett [1962] and Woolf [1955].

While Good's and Goodman's formulations and tests of no-interaction hypotheses are entirely general, physical interpretations of their meanings became extremely difficult, if not impossible, for interactions higher than the second, in which case these interactions reduce to the ones discussed before. Bhapkar and Koch [1965, 1966, 1968] outlined the models, tests, and interpretations of the hypothesis of no interaction in three-dimensional and four-dimensional contingency tables in great detail, and compared results of using different statistics, all based on Wald's criterion, for several numerical examples.

1.5. Conditions Essential to a Definition of Hypothesis of No Interaction in Multidimensional Contingency Tables

Based on the above review of treatments of higher-order interactions, it appears that several basic and related concepts are important in the formulation of its definition:

a. The Fixed Marginal Totals

In fact, to talk about the r th-order interaction in an m -way table, the r -way marginals must be considered fixed; for otherwise we would be considering a less restricted hypothesis which includes the no-interaction hypothesis as a subhypothesis. This concept was implied in Bartlett's and Roy's definition, and explicitly stated in Good's definition. Darroch and Birch also assumed fixed marginals. Goodman and Bhapkar, on the other hand, did not make such a demand on their definitions, and hence their formulations are less desirable in the sense that their interaction *does not measure the interaction as given by the data*, but that possibly for another set of data with somewhat different marginal totals.

b. The Requirements of Symmetry

This requirement is a "logically attractive" condition as stated by Simpson, and demands that the statistic be invariant upon relabeling of the classifications. This requirement is again satisfied by all investigations with the exception of Goodman, and Bhapkar and Koch, who maintained that the symmetry requirement is not necessarily desirable for certain physical interpretations.

c. Unique Set of Cell Probabilities ($p_i > 0$)

The no-interaction hypothesis, which is presumably the last hypothesis to be tested in a hierarchy of hypotheses, should determine the cell probabilities uniquely. This condition was conjectured by Darroch for his formulation of the no second-order interaction hypothesis and shown to be true by Birch and Good. A measure of deviations of the data from this set of cell probabilities would therefore be a measure of interaction.

d. Additivity of the Statistics

The "mechanism" used by Roy and Kastenbaum in their definition of no interaction demands that H_1 (R is independent of C) \cap H_2 (R is independent of D) \cap H_1 (no interaction) implies H (R is independent of CD). The same requirement was also discussed by Birch [1963]. In general, if a more restrictive hypothesis can be considered as the intersection of several less restrictive hypotheses, it would be logical and desirable to require the test statistics for the component hypotheses to sum up to that of the more restrictive hypothesis. This requirement is not fulfilled by the usual X^2 statistic except in an asymptotic sense. The additive analysis of component variation, similar to that of analysis of variance, is a desirable feature of information analysis as noted by Lewis [1962].

2. Interaction from the Viewpoint of Information Theory

2.1. The Minimum Discrimination Information Statistic (m.d.i.s.)

In the analysis of contingency table data, two major types of hypotheses are usually postulated and tested. One is a test of the cell frequencies of an observed sample table against known or given probabilities of a table of the same size, the other is a test of the structural relationships that seem reasonable among the classifications, e.g., association or independence among responses, homogeneity of response over several categories of a factor, etc.

In Kullback [1959], and two papers by Kullback, Kupperman, and Ku [1962a, 1962b], a number of useful tests for contingency tables based on the notion of information theory are given for hypotheses which can be expressed explicitly as functions of specified marginal probabilities. The minimum discrimination information statistic, m.d.i.s., was suggested as the test statistic which, in its simplest form for a two-way contingency table, can be expressed as

$$(2.1) \quad 2n\hat{I}(p:\pi) = 2\sum_{ij} x_{ij} \ln \frac{x_{ij}}{n\pi_{ij}},$$

where π_{ij} is the probability of an observation from the i th row and j th column of the table under the null hypothesis, $\sum_{ij} \pi_{ij} = 1$, x_{ij} is the observed frequency of occurrence in the corresponding cell, $\sum_{ij} x_{ij} = n$, and \ln is the natural logarithm. $0 \ln 0$ is defined as zero.

Similarly, the m.d.i.s. for the test of independence between the row and column classifications, i.e., $\pi_{ij} = \pi_{i.}\pi_{.j}$, is shown to be

$$(2.2) \quad 2n\hat{I}(p:\pi) = 2\sum_{ij} x_{ij} \ln \frac{x_{ij}}{\frac{x_{i.}x_{.j}}{n}},$$

where $x_{i.} = \sum_j x_{ij}$, $x_{.j} = \sum_i x_{ij}$, are the row and column marginal frequencies of the two-way table.

We may consider either of the two tests given by (2.1) and (2.2) to be a comparison of the observed frequencies against a set of frequencies in a constructed table represented by $x_{ij}^* = np_{ij}^*$, where $\{p_{ij}^*\}$ is the set of cell probabilities that "most" resembles $\{\pi_{ij}\}$ subject to certain marginal probability restrictions. In fact, the set of $\{p_{ij}^*\}$ can be obtained by minimizing the discrimination information

$$(2.3) \quad 2I(p:\pi) = 2\sum_{ij} p_{ij} \ln \frac{p_{ij}}{\pi_{ij}}$$

subject to these restrictions. For the case considered under (2.1), the minimum value is zero for $p_{ij}^* = \pi_{ij}$, the restriction $\sum_{ij} np_{ij}^* = n$ is always fulfilled. For the case (2.2), the minimum value of (2.3) is attained for $p_{ij}^* = p_{i.}p_{.j}$, with the restriction $np_{i.} = x_{i.}$, and $np_{.j} = x_{.j}$, as shown in Kullback et al. [1962b].

It would clearly be desirable if this concept could be extended to the formulation of second-order and higher-order interactions. Here, however, we encounter essentially the same difficulties as discussed by Darroch, i.e., these interaction hypotheses cannot be formulated in terms of explicit functions of the marginal probabilities such that these functions also satisfy all the fixed marginal total restraints. To resolve these difficulties, we need to give a number of new results with information theoretical background due to Ireland and Kullback [1968] in conjunction with their study of an estimation problem first considered by Deming and Stephan [1940]. We shall summarize these results in the next section and show how they may be applied to give a unified approach to the analysis of interactions in multidimensional contingency tables.

2.2. Summary of Current Results

In the following we shall present a number of current results in the form of three theorems. Proofs for theorems 2.1 and 2.2 using properties of the minimum discrimination information are given in Ireland and Kullback [1968], and will not be repeated here. The use and interpretations of these theorems in our case for the purpose of hypothesis testing, however, are quite different from that of their paper which essentially treats a problem in estimation. These differences will be discussed and appropriate modifications to the statement of these theorems incorporated.

THEOREM 2.1. *Given a contingency table $\{\pi_{ij}\}$, $i=1, 2, \dots, r$, $j=1, \dots, c$, $\pi_{ij} > 0$, $\sum_{ij} \pi_{ij} = 1$. Consider all contingency tables $\{p_{ij}\}$ of the same dimension such that the marginal probabilities $p_{i.} = \sum_j p_{ij}$ and $p_{.j} = \sum_i p_{ij}$ are given and fixed. Then the minimum value of the discrimination information*

$$(2.4) \quad I(p:\pi) = \sum_{ij} p_{ij} \ln \frac{p_{ij}}{\pi_{ij}}$$

is attained for $p_{ij} = p_{ij}^ = a_i b_j \pi_{ij}$ where the a_i 's and b_j 's are determined subject to the marginal probability restrictions.*

Deming and Stephan [1940] considered the problem of estimation of cell probabilities from a sample of observations in an $r \times c$ table for which the population marginal probabilities $p_{i.}$ and $p_{.j}$ are known and fixed. Hence if we use the maximum likelihood estimates of the cell probabilities, $\pi_{ij} = n_{ij}/n$, which do not necessarily satisfy these marginal restraints, then the question can be posed "What is the p_{ij} distribution satisfying these restraints and also 'closest' to the observed sample in some sense?" Deming and Stephan [1940] suggested using estimates that minimized

$$(2.5) \quad (X')^2 = \sum_{ij} (n_{ij} - np_{ij})^2 / n_{ij}.$$

Ireland and Kullback [1968] suggested minimizing the discrimination information (2.4) which generates RBAN estimators, as does Deming and Stephan's procedure.

We note that π_{ij} is not specified in the theorem. If, instead of letting $\pi_{ij} = n_{ij}/n$, we use π_{ij} to represent the cell probabilities of some reasonable hypothesis which we are interested in, or the ausgezeichnet hypothesis in the sense of Good, then the p_{ij}^* distribution will represent the distribution that is "closest" to this hypothesized distribution subject to the marginal restraints in the sense of minimum discrimination information or "minimal discriminability". For instance, if $\pi_{ij} = \pi_{i.} \pi_{.j}$, or the hypothesis of independence of row and column classifications, then by theorem 2.1,

$$\begin{aligned} p_{ij}^* &= a_i b_j \pi_{ij}, \\ p_{i.} &= \sum_j p_{ij}^* = a_i \sum_j b_j \pi_{ij}, \\ p_{.j} &= \sum_i p_{ij}^* = b_j \sum_i a_i \pi_{ij}, \end{aligned}$$

and

$$(2.6) \quad p_{ij}^* = \frac{p_{i.}}{\pi_{i.} \sum_j b_j \pi_{.j}} \cdot \frac{p_{.j}}{\pi_{.j} \sum_i a_i \pi_{i.}} \cdot \pi_{ij} = \frac{p_{i.} p_{.j}}{(\sum_j b_j \pi_{.j}) (\sum_i a_i \pi_{i.})} = p_{i.} p_{.j}$$

since $\sum_{ij} a_i b_j \pi_{ij} = 1$.

The hypothesis to be tested is then the independence of the two classifications subject to the restraints $np_{i.} = x_{i.}$ and $np_{.j} = x_{.j}$, the fixed marginal totals.

Theorem 2.1 is stated in terms of a two-way contingency table for notational convenience. For a three-way contingency table, if $\pi_{ijk} = \pi_{i..} \pi_{.j.} \pi_{..k}$, then $p_{ijk}^* = a_i b_j c_k \pi_{ijk}$ and the hypothesis to be

tested is $p_{ijk} = p_{i..}p_{.j.}p_{..k}$, following the same derivation as the two-way table exactly. We shall define all such hypotheses where the one-way marginals are completely specified as the no first-order interaction hypothesis.

Extending this concept, we shall define no second-order interaction in a three-way contingency table as represented by the p_{ijk}^* distribution when all the two-way marginals are considered as fixed and for appropriate selection of π_{ijk} . The justification for this formulation will be discussed in subsection 3.2. Theorem 2.1 can be then restated as

THEOREM 2.1A. *Given a contingency table $\{\pi_{ijk}\}$, $i = 1, 2, \dots, r$, $j = 1, 2, \dots, c$, $k = 1, 2, \dots, d$, $\sum_{ijk} \pi_{ijk} = 1$, where the $\pi_{ijk} > 0$ represent cell probabilities of some reasonable hypothesis. Consider all contingency tables $\{p_{ijk}\}$ of the same dimension with fixed two-way marginals $p_{i..}$, $p_{.jk}$, and $p_{i..k}$, then the minimum value of the quantity*

$$I(p : \pi) = \sum_{ijk} p_{ijk} \ln \frac{p_{ijk}}{\pi_{ijk}}$$

is attained for $p_{ijk} = p_{ijk}^ = a_{ij}b_{jk}c_{ik}\pi_{ijk}$, $\sum_{ijk} a_{ij}b_{jk}c_{ik}\pi_{ijk} = 1$, where a_{ij} , b_{jk} , and c_{ik} are functions of the given two-way marginal probabilities. Equivalently, the condition may be stated as*

$$(2.7) \quad \ln \frac{p_{ijk}^*}{\pi_{ijk}} = \ln a_{ij} + \ln b_{jk} + \ln c_{ik},$$

representing no second-order interaction among the three classifications.

To compute the numerical values of p_{ijk}^* , we need

THEOREM 2.2. *The set of p_{ij}^* in Theorem 2.1 can be computed by an iterative procedure alternatively satisfying one and then the other marginal restraints. The iteration is given by*

$$p_{ij}^{(2n-1)} = \frac{p_{i.}}{p_{i.}^{(2n-2)}} p_{ij}^{(2n-2)}, \quad p_{ij}^{(2n)} = \frac{p_{.j}}{p_{.j}^{(2n-1)}} p_{ij}^{(2n-1)}, \quad n = 1, 2, \dots, \quad p_{ij}^{(0)} = \pi_{ij}.$$

If $p_{ij}^{(N)}$ represent the value of the cell probabilities after the Nth iteration, then either $p_{ij}^{(N)} = f_{ij} = p_{ij}^$ for some finite N, or $p_{ij}^{(\infty)} = p_{ij}^*$.*

We shall indicate the first few steps of the iteration process for a two-way table using the relationship

$$p_{ij}^* = a_i b_j \pi_{ij}, \quad p_{i.} = a_i \sum_j b_j \pi_{ij}, \quad p_{.j} = b_j \sum_i a_i \pi_{ij}$$

Let $b_j^{(1)} = 1$, then

$$p_{i.} = a_i^{(1)} \pi_{i.}, \quad p_{ij}^{(1)} = \frac{p_{i.}}{\pi_{i.}} \pi_{ij} = a_i^{(1)} b_j^{(1)} \pi_{ij},$$

$$p_{.j} = b_j^{(2)} \sum_i a_i^{(1)} \pi_{ij}, \quad p_{ij}^{(2)} = \frac{p_{.j}}{p_{.j}^{(1)}} p_{ij}^{(1)} = a_i^{(1)} b_j^{(2)} \pi_{ij},$$

$$p_{i.} = a_i^{(2)} \sum_j b_j^{(2)} \pi_{ij}, \quad p_{ij}^{(3)} = \frac{p_{i.}}{p_{i.}^{(2)}} p_{ij}^{(2)} = a_i^{(2)} b_j^{(2)} \pi_{ij},$$

$$p_{.j} = b_j^{(3)} \sum_i a_i^{(2)} \pi_{ij}, \quad p_{ij}^{(4)} = \frac{p_{.j}}{p_{.j}^{(3)}} p_{ij}^{(3)} = a_i^{(2)} b_j^{(3)} \pi_{ij},$$

.....

$$p_{i\cdot} = a_i^{(n)} \sum_j b_j^{(n)} \pi_{ij} = p_{i\cdot}^{(2n-1)}, \quad p_{ij}^{(2n-1)} = \frac{p_{i\cdot}}{p_{i\cdot}^{(2n-2)}} p_{ij}^{(2n-2)} = a_i^{(n)} b_j^{(n)} \pi_{ij},$$

$$p_{\cdot j} = b_j^{(n+1)} \sum_i a_i^{(n)} \pi_{ij} = p_{\cdot j}^{(2n)}, \quad p_{ij}^{(2n)} = \frac{p_{\cdot j}}{p_{\cdot j}^{(2n-1)}} p_{ij}^{(2n-1)} = a_i^{(n)} b_j^{(n+1)} \pi_{ij},$$

etc.

Of course if $\pi_{ij} = \pi_{i\cdot} \pi_{\cdot j}$, then after one cycle of iteration on the two marginals, we have

$$\begin{aligned} p_{ij}^{(2)} &= \frac{p_{\cdot j}}{p_{\cdot j}^{(1)}} p_{ij}^{(1)} = \frac{p_{\cdot j}}{\sum_i \frac{p_{i\cdot}}{\pi_{i\cdot}} \pi_{i\cdot} \pi_{\cdot j}} \cdot \frac{p_{i\cdot}}{\pi_{i\cdot}} \pi_{i\cdot} \pi_{\cdot j} \\ &= \frac{p_{\cdot j}}{\pi_{\cdot j}} \cdot p_{i\cdot} \pi_{\cdot j} \\ &= p_{i\cdot} p_{\cdot j} \end{aligned}$$

and the marginal restraints are satisfied exactly. Hence the iteration process terminates and the solution is exact.

The results of iteration with different sets of specified marginals for a four-way contingency table π_{ijkl} , $i = 1, \dots, r$, $j = 1, \dots, c$, $k = 1, \dots, d$, and $l = 1, \dots, t$, are given below for purpose of illustration.

Let the given marginals be

$$(2.8) \quad p_{i\cdot\cdot\cdot}, \quad p_{\cdot j\cdot\cdot}, \quad p_{\cdot\cdot k\cdot}, \quad p_{\cdot\cdot\cdot l}$$

then

$$\begin{aligned} (2.9) \quad p_{ijkl}^* &= a_i b_j c_k d_l \pi_{ijkl} \\ p_{i\cdot\cdot\cdot} &= a_i \sum_{jkl} b_j c_k d_l \pi_{ijkl} \\ p_{\cdot j\cdot\cdot} &= b_j \sum_{ikl} a_i c_k d_l \pi_{ijkl} \\ p_{\cdot\cdot k\cdot} &= c_k \sum_{ijl} a_i b_j d_l \pi_{ijkl} \\ p_{\cdot\cdot\cdot l} &= d_l \sum_{ijk} a_i b_j c_k \pi_{ijkl} \end{aligned}$$

The iterative solution of the system (2.9) cycles through

$$\begin{aligned} (2.10) \quad p_{ijkl}^{(4n+1)} &= \frac{p_{i\cdot\cdot\cdot}}{p_{i\cdot\cdot\cdot}^{(4n)}} p_{ijkl}^{(4n)}, & p_{ijkl}^{(4n+2)} &= \frac{p_{\cdot j\cdot\cdot}}{p_{\cdot j\cdot\cdot}^{(4n+1)}} p_{ijkl}^{(4n+1)}, \\ p_{ijkl}^{(4n+3)} &= \frac{p_{\cdot\cdot k\cdot}}{p_{\cdot\cdot k\cdot}^{(4n+2)}} p_{ijkl}^{(4n+2)}, & p_{ijkl}^{(4n+4)} &= \frac{p_{\cdot\cdot\cdot l}}{p_{\cdot\cdot\cdot l}^{(4n+3)}} p_{ijkl}^{(4n+3)}. \end{aligned}$$

Let the given marginals be

$$(2.11) \quad p_{i\cdot\cdot\cdot}, \quad p_{\cdot j\cdot\cdot}, \quad \text{and } p_{\cdot\cdot kl}.$$

Then the problem is essentially that of a three-way table p_{ijm} where $m = (k, l) = 1, \dots, dt$. Since given $p_{..kl}$ determines the marginals $p_{..k.}$ and $p_{...l}$, that is (2.11) \Rightarrow (2.8),

$$\sum_{ijk} p_{ijk}^* \ln \frac{p_{ijk}^*}{\pi_{ijk}} \leq \sum_{ijm} p_{ijm}^* \ln \frac{p_{ijm}^*}{\pi_{ijm}}.$$

Let the given marginals be

$$(2.12) \quad p_{ij..}, \quad p_{i..k}, \quad p_{i..l}, \quad p_{.jk.}, \quad p_{.j.l}, \quad p_{..kl},$$

then

$$\begin{aligned}
(2.13) \quad & p_{ijkl}^* = a_{ij} b_{ik} c_{il} d_{jk} e_j f_{kl} \pi_{ijk l}, \\
& p_{ij..} = a_{ij} \sum_k b_{ik} c_{il} d_{jk} e_j f_{kl} \pi_{ijk l}, \\
& . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \\
& p_{..kl} = f_{kl} \sum_{ij} a_{ij} b_{ik} c_{jl} d_{jk} e_{il} \pi_{ijk l},
\end{aligned}$$

and the iterative solution of the system (2.13) cycles through

$$(2.14) \quad p_{ijkl}^{(6n+1)} = \frac{p_{ij\cdot\cdot}}{p_{ii\cdot\cdot}} p_{ijkl}^{(6n)}, \quad \cdot \cdot \cdot, \quad p_{ijkl}^{(6n+6)} = \frac{p_{\cdot\cdot kl}}{p_{\cdot\cdot kl}^{(6n+5)}} p_{ijkl}^{(6n+5)}.$$

Since the marginals given in (2.12) determine those given in (2.8) and (2.11), i.e.,

(2.12) \Rightarrow (2.11) \Rightarrow (2.8), it follows that

$$(2.15) \quad I(p^* : \pi|2.8) \leq I(p^* : \pi|2.11) \leq I(p^* : \pi|2.12).$$

If the given marginals are

$$(2.16) \quad p_{ijk\cdot}, \quad p_{ij\cdot l}, \quad p_{i\cdot kl} \text{ and } p_{\cdot jkl},$$

then

$$\begin{aligned}
 (2.17) \quad & p_{ijkl}^* = a_{ijk} b_{ijl} c_{ikl} d_{jkl} \pi_{ijkl} \\
 & p_{ijk.} = a_{ijk} \sum_l b_{ijl} c_{ikl} d_{jkl} \pi_{ijkl}, \\
 & \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
 & p_{.jkl} = d_{jkl} \sum_i a_{ijk} b_{ijl} c_{ikl} \pi_{ijkl}.
 \end{aligned}$$

and the iterative solution of the system (2.17) cycles through

$$p_{ijkl}^{(4n+1)} = \frac{p_{ijk} \cdot p_{jkl}}{p_{ijk}^{(4n)}} p_{ijkl}^{(4n)}, \quad \dots, \quad p_{ijkl}^{(4n+4)} = \frac{p \cdot jkl}{p_{jkl}^{(4n+3)}} p_{ijkl}^{(4n+3)}.$$

Since (2.16) \Rightarrow (2.12) \Rightarrow (2.11) \Rightarrow (2.8), it follows that

$$(2.18) \quad I(p^*: \pi|2.8) \leq I(p^*: \pi|2.11) \leq I(p^*: \pi|2.12) \leq I(p^*: \pi|2.16).$$

These relationships will be useful in the construction of analysis of information tables to be described in subsections 3.1 and 3.2. The appropriate choice of π_{ijkl} will also be deferred for later discussion.

A third result we shall need is

THEOREM 2.3 *The equality*

$$(2.19) \quad I(p : \pi) = I(p : p^*) + I(p^* : \pi)$$

holds true for p^* computed by the procedure stated in Theorem 2.2 where the p^* - and p -distribution have common specified marginals.

This is a special case of a property of the minimum discrimination information and can be deduced from a theorem in Kullback and Khairat [1966]. We demonstrate the theorem as applied to a three-way contingency table when all two-way marginals are considered fixed, and where

$$\sum_k p_{ijk}^* = \sum_k p_{ijk} = p_{ij\cdot},$$

$$\sum_j p_{ijk}^* = \sum_j p_{ijk} = p_{i\cdot k},$$

$$\sum_i p_{ijk}^* = \sum_i p_{ijk} = p_{\cdot jk}.$$

We have

$$\begin{aligned} (2.20) \quad & \sum_{ijk} p_{ijk}^* \ln \frac{p_{ijk}^*}{\pi_{ijk}} \\ &= \sum_{ijk} p_{ijk}^* \ln a_{ij} + \sum_{ijk} p_{ijk}^* \ln b_{jk} + \sum_{ijk} p_{ijk}^* \ln c_{ik} \\ &= \sum_{ij} p_{ij\cdot} \ln a_{ij} + \sum_{jk} p_{\cdot jk} \ln b_{jk} + \sum_{ik} p_{i\cdot k} \ln c_{ik} \\ &= \sum_{ijk} p_{ijk} \ln \frac{p_{ijk}^*}{\pi_{ijk}}. \end{aligned}$$

Hence,

$$\begin{aligned} I(p : \pi) &= \sum_{ijk} p_{ijk} \ln \frac{p_{ijk}}{\pi_{ijk}} \\ &= \sum_{ijk} p_{ijk} \ln \frac{p_{ijk}}{p_{ijk}^*} + \sum_{ijk} p_{ijk} \ln \frac{p_{ijk}^*}{\pi_{ijk}} \\ &= I(p : p^*) + I(p^* : \pi). \end{aligned}$$

2.3. The No-Interaction Hypothesis as a Form of "Generalized Independence"

The formulation of no second-order interaction given in theorem 2.1A suggests that all no interaction hypotheses can be defined in a similar manner, depending on the marginals which are considered given and fixed. Since higher-order marginals determine all lower-order marginals, it is natural to consider the π -distribution as the uniform distribution, corresponding to the case where no marginals are specified, as a general form of independence. For the uniform distribution,

$p_{ijkl}^* = \pi_{ijkl} = \frac{1}{rcdt}$ in a four-way contingency table. Given all one-way marginals and taking π_{ijkl}

$= \frac{1}{rcdt}$, we obtain

$$(2.21) \quad p_{ijkl}^* = p_{i...} p_{.j..} p_{..k.} p_{...l},$$

i.e., the distribution which “most resembles” the uniform distribution subject to the four one-way marginal restraints. Derivation of (2.21) follows exactly as that of (2.6). Given all two-way marginals, the distribution that “most resembles” the uniform distribution subject to the six two-way marginals restraints is the p_2^* distribution representing no second-order interaction. Given all four three-way marginals, we obtain p_3^* corresponding to no third-order interaction.

In this sense the no-interaction hypotheses can be considered as generalized forms of independence hypotheses, where the “degrees” of independence that can be realized depend on the marginal restraints imposed. Each time we add on a restraint, we obtain a p^* distribution corresponding to the condition of minimum discrimination information subject to the additional restraints, and corresponding to the appropriate null hypothesis to be tested given this additional restraint.

Hence we may state the principle of minimum discrimination information for the generation of appropriate hypotheses:

“If certain marginal probabilities of a contingency table are considered given or fixed, then the appropriate interaction hypothesis to be tested, subject to these fixed marginal restraints, is the hypothesis represented by the unique set of cell probabilities p_{ijkl}^* satisfying these restraints and yielding the minimum value of discrimination information

$$(2.22) \quad I(p : \pi) = \sum_{ijkl} p_{ijkl} \ln p_{ijkl} + \ln rcdt$$

for all p_{ijkl} .”

It can be shown that all the usual “classical” hypotheses can be generated by the application of this principle. If complete sets of marginals are considered given in a four-way table, we arrive at the following sequence.

Marginals considered as fixed		No-interaction hypothesis
$p_{....} = 1$	$p_0^* = \frac{1}{rcdt}$	zeroth-order (uniform).
$p_{i...}, \quad p_{.j..}, \quad p_{..k.}, \quad p_{...l}$	p_1^*	first-order (4-way independence).
$p_{ij..}, \quad p_{i.k.}, \quad p_{i..l}, \quad p_{.jk.}, \quad p_{.j.l}, \quad p_{..kl}$	p_2^*	second-order.
$p_{ijk.}, \quad p_{ij.l}, \quad p_{i.kl}, \quad p_{.jkl}$	p_3^*	third-order.
p_{ijkl}	$p_4^* = p_{ijkl}$	fourth-order (no test).

If only part of a complete set of marginals are given, a conditional type of independence is generated. Some of these hypotheses for which the p_{ijkl}^* can be explicitly expressed in terms of the marginals are given in table 2.1. We demonstrate here the generation of the conditional inde-

pendence in a three-way table when two of the two-way marginals, $p_{ij\cdot}$ and $p_{\cdot jk}$ are assumed to be fixed.

TABLE 2.1. *Some explicit expressions for p_{ijk}^**

Marginal restraints	p_{ijk}^*
$p_{\cdot\cdot\cdot\cdot} = 1$	$1/rcdt$
$p_{i\cdot\cdot\cdot}$	$p_{i\cdot\cdot\cdot}/cdt$
$p_{i\cdot\cdot\cdot}, p_{\cdot j\cdot\cdot}$	$p_{i\cdot\cdot\cdot}p_{\cdot j\cdot\cdot}/dt$
$p_{i\cdot\cdot\cdot}, p_{\cdot j\cdot\cdot}, p_{\cdot\cdot k\cdot}$	$p_{i\cdot\cdot\cdot}p_{\cdot j\cdot\cdot}p_{\cdot\cdot k\cdot}/t$
$p_{i\cdot\cdot\cdot}, p_{\cdot j\cdot\cdot}, p_{\cdot\cdot k\cdot}, p_{\cdot\cdot\cdot l}$	$p_{i\cdot\cdot\cdot}p_{\cdot j\cdot\cdot}p_{\cdot\cdot k\cdot}p_{\cdot\cdot\cdot l}$
$p_{ij\cdot\cdot}$	$p_{ij\cdot\cdot}/dt$
$p_{ij\cdot\cdot}, p_{\cdot\cdot kl}$	$p_{ij\cdot\cdot}p_{\cdot\cdot kl}$
$p_{\cdot j\cdot\cdot}, p_{i\cdot\cdot l}, p_{\cdot\cdot kl}$	$p_{\cdot j\cdot\cdot}p_{i\cdot\cdot l}p_{\cdot\cdot kl}/p_{\cdot\cdot\cdot l}$
$p_{ij\cdot\cdot}, p_{i\cdot k\cdot}, p_{i\cdot\cdot l}$	$p_{ij\cdot\cdot}p_{i\cdot k\cdot}p_{i\cdot\cdot l}/(p_{i\cdot\cdot\cdot})^2$
$p_{ij\cdot\cdot}, p_{i\cdot\cdot l}, p_{\cdot\cdot kl}$	$p_{ij\cdot\cdot}p_{i\cdot\cdot l}p_{\cdot\cdot kl}/p_{i\cdot\cdot\cdot}p_{\cdot\cdot\cdot l}$
$p_{ij\cdot\cdot}, p_{i\cdot\cdot l}, p_{\cdot jk\cdot}$	no explicit expression

By theorem 2.1A we have

$$p_{ijk}^* = a_{ij}b_{jk}\pi_{ijk},$$

subject to the restraints

$$\sum_i p_{ijk}^* = b_{jk} \sum_i a_{ij} \pi_{ijk} = p_{\cdot jk},$$

$$\sum_k p_{ijk}^* = a_{ij} \sum_k b_{jk} \pi_{ijk} = p_{ij\cdot}.$$

Hence

$$(2.23) \quad p_{ijk}^* = \frac{p_{\cdot jk}}{(\sum_i a_{ij} \pi_{ijk})} \cdot \frac{p_{ij\cdot}}{(\sum_k b_{jk} \pi_{ijk})} \cdot \pi_{ijk}.$$

If we let $\pi_{ijk} = \frac{1}{rcd}$, then

$$(2.24) \quad p_{ijk}^* = \frac{p_{\cdot jk} p_{ij\cdot}}{\sum_{ik} a_{ij} b_{jk} \pi_{ijk}} = \frac{p_{\cdot jk} p_{ij\cdot}}{p_{\cdot j\cdot}},$$

or the conditional independence hypothesis of the row and depth classification given the column classification.

It is of interest to note that the same expression for p_{ijk}^* in (2.24) is obtained whether we use

$$(2.25) \quad \pi_{ijk} = 1/rcd$$

$$\pi_{ijk} = \pi_{i..} \pi_{.j.} \pi_{..k}, \text{ or}$$

$$\pi_{ijk} = \pi_{ij.} \pi_{.jk} / \pi_{.j.}$$

Thus, as long as π_{ijk} represent the condition of generalized independence corresponding to marginal restraints of an order lower or equal to the given marginal restraints, the same set of p_{ijk}^* will be generated by our procedure. In the present case

$$\text{given} \quad (p_{ij.}, p_{.jk}) \Rightarrow (p_{i..}, p_{.j.}, p_{..k}) \Rightarrow p_{...}$$

A simple proof of this property is as follows. The set of p_{ijk}^* is obtained by minimizing the discrimination information

$$I(p : \pi) = \sum_{ijk} p_{ijk} \ln \frac{p_{ijk}}{\pi_{ijk}}$$

subject to given marginal restraints. Now

$$(2.26) \quad I(p : \pi) = \sum_{ijk} p_{ijk} \ln p_{ijk} - \sum_{ijk} p_{ijk} \ln \pi_{ijk}$$

and the second term of the expression in the right-hand side of (2.26) reduces to a constant no matter which form of π_{ijk} in (2.25) is used. Hence the same p_{ijk}^* will give the minimum value of (2.26). Good [1966] showed that the chains of hypotheses generated by the principle of minimum discriminability depend only on the increasing sequence of linear constraints, irrespective of which of the existing hypotheses the new ones are referred to.

2.4. Consistency of Information-Theoretic Definition of No Second-Order Interaction With Other Formulations

We shall show that the definition of no second-order interaction given in Theorem 2.1A is consistent with the formulations given by Bartlett, and Roy and Kastenbaum represented by (1.1) and (1.2). We may remark here also that the p^* -distribution satisfies all the four requirements in subsection 1.4. The p^* -distribution satisfies the requirements of (a) fixed marginal totals (b) symmetry and (c) unique set of cell probabilities because of the way it is derived. Additivity is a property of m.d.i.s. [Theorem 2.1, ch. 2, Kullback 1959] which facilitates the construction of analysis of information tables.

For an $r \times c$ table with given set of cell probabilities p_{ij} and π_{ij} , let us find the table with the same marginals as p_{ij} but minimizing the expression

$$f(d_{ij}) = \sum_{ij} (p_{ij} + d_{ij}) \ln \frac{(p_{ij} + d_{ij})}{\pi_{ij}}$$

with

$$\sum_i d_{ij} = \sum_j d_{ij} = 0.$$

Since $f(d_{ij})$ is a convex function, its minimum is given by the d_{ij} 's satisfying the set of equations

$$\frac{\partial f(d_{ij})}{\partial d_{ij}} = 0 = \ln \frac{p_{ij} + d_{ij}}{\pi_{ij}} - \ln \frac{p_{ic} + d_{ic}}{\pi_{ic}} - \ln \frac{p_{rj} + d_{rj}}{\pi_{rj}} + \ln \frac{p_{rc} + d_{rc}}{\pi_{rc}}.$$

For the values of d_{ij} 's satisfying the above let $p_{ij} + d_{ij} = p_{ij}^*$, then the set of equations reduces to

$$(2.27) \quad \frac{p_{ij}^*}{\pi_{ij}} \cdot \frac{p_{rc}^*}{\pi_{rc}} = \frac{p_{ic}^*}{\pi_{ic}} \cdot \frac{p_{rj}^*}{\pi_{rj}},$$

and all are satisfied by

$$p_{ij}^* = a_i b_j \pi_{ij}.$$

This procedure is essentially what Bartlett used in getting a solution to the no second-order interaction hypothesis in a $2 \times 2 \times 2$ table. Bartlett specified the condition of no second-order interaction to be

$$(2.28) \quad \frac{p_{111}p_{122}}{p_{112}p_{121}} = \frac{p_{211}p_{222}}{p_{212}p_{221}}.$$

For observed cell frequencies x_{ijk} , he solved for Δ in the equation

$$(2.29) \quad \frac{(x_{111} + \Delta)(x_{122} + \Delta)}{(x_{112} - \Delta)(x_{121} - \Delta)} = \frac{(x_{211} - \Delta)(x_{222} - \Delta)}{(x_{212} + \Delta)(x_{221} + \Delta)},$$

and computed

$$X^2 = \Delta^2 \sum_{ijk} \frac{1}{x_{ijk}}.$$

In fact, Bartlett's X^2 is an approximation to $2n\hat{I}$ given in (3.3). If we let $x_{ijk} = x_{ijk}^* \pm \Delta$ and expand $x_{ijk} \ln x_{ijk}$ about x_{ijk}^* in x_{ijk}^* by a Taylor series expansion up to Δ^2 , we have

$$x_{ijk} \ln x_{ijk} \doteq x_{ijk}^* \ln x_{ijk}^* \pm \Delta (\ln x_{ijk}^* + 1) + \frac{\Delta^2}{2} \left(\frac{1}{x_{ijk}^*} \right).$$

Summing over $i, j, k = 1, 2$,

$$\sum_{ijk} x_{ijk} \ln x_{ijk} \doteq \sum_{ijk} x_{ijk}^* \ln x_{ijk}^* + \sum_{ijk} (\pm \Delta) (\ln x_{ijk}^* + 1) + \frac{\Delta^2}{2} \sum_{ijk} \frac{1}{x_{ijk}^*}.$$

Since the middle term on the right hand side is equal to zero by (2.28),

$$2n\hat{I} = 2 \sum_{ijk} x_{ijk} \ln \frac{x_{ijk}}{x_{ijk}^*} \doteq \Delta^2 \sum_{ijk} \frac{1}{x_{ijk}^*} = X^2.$$

Also it can be checked readily by substitution that if a π_{ijk} distribution satisfies (2.28), then the p_{ijk}^* distribution also satisfies (2.28) when $p_{ijk}^* = a_i b_j c_k \pi_{ijk}$. The same is true for the set of conditions of no second-order interaction in an $r \times c \times d$ table given by Roy and Kastenbaum in eq (1.2).

It is to be noted also that Good's Principle of Minimal Discriminability is essentially the same as our Principle of Minimum Discrimination Information. Darroch's suggested procedure and our

procedure also share a number of similarities. However, we have demonstrated that many hypotheses of interest for contingency tables can be generated through one unified procedure based on information theory and the hypothesis of no second-order interaction is no exception. Furthermore, the properties of the minimum discrimination information statistics, convexity and additivity, allow us to derive and interpret these statistics naturally, and to construct the analysis of information tables for an overall comparison between the sample and the nested sequence of hypotheses.

For other related literature dealing with a similar problem but not in the testing of no-interaction hypotheses in contingency tables, we cite P. M. Lewis II [1959] and D. T. Brown [1959]. Lewis, on an *ad hoc* basis, considered approximations to a discrete probability distribution over the space of binary random variables by distributions which are products of various marginal distributions of the original one and measured the goodness of fit by the discrimination information measure. D. T. Brown extended this notion to using an approximating distribution having certain marginals the same as those of the original distribution. He described an iterative procedure which is the same as the example following theorem 2.2, and showed that the goodness of the approximation improves at each step of the iteration using discrimination information as a measure of goodness. The procedure he described is the special case of an n -way $2 \times 2 \times \dots \times 2$ table with initial cell probabilities all equal to $1/2^n$.

3. Analysis of Multidimensional Contingency Tables and the Interpretation of Results

3.1. The m.d.i.s. for the No-Interaction Hypothesis

In the last section we have shown that the no-interaction hypothesis can be considered as a hypothesis of generalized independence, subject to fixed marginal restraints. Furthermore, the unique set of p^* -distributions can be computed by a convergent iterative process alternatively satisfying the given marginals.

Let p_0 represent the cell probabilities under the hypothesis of uniform distribution. We say that $\{p^*\}$ is the table that most resembles $\{p_0\}$ subject to these marginal restraints, or there is no interaction between $\{p^*\}$ and $\{p_0\}$ in the sense that

$$I(p^*:p_0) = \sum p^* \ln \frac{p^*}{p_0}$$

is a minimum for all p 's consistent with dimension of the table and the given restraints. For any other table we may write

$$(3.1) \quad \sum p \ln \frac{p}{p_0} = \sum p^* \ln \frac{p^*}{p_0} + \left\{ \sum p \ln \frac{p}{p_0} - \sum p^* \ln \frac{p^*}{p_0} \right\}.$$

Since the first term on the right represents the condition of no interaction, the term in the bracket

$$(3.2) \quad \sum p \ln \frac{p}{p_0} - \sum p^* \ln \frac{p^*}{p_0} = \sum p \ln \frac{p}{p^*}$$

is then a measure of interaction, or the departure of the p -distribution from the no-interaction distribution. The equality in (3.2) follows from theorem 2.3.

Given an observed sample with cell frequencies x_{ijk} in a three-way table, and let $x_{ij.}$, $x_{i.k}$, and $x_{.jk}$ be the given marginal restraints, then the m.d.i.s. for testing the hypothesis of no second-order interaction is

$$(3.3) \quad 2nI(\hat{p}:p^*) = 2\sum_{ijk} x_{ijk} \ln \frac{x_{ijk}}{x_{ijk}^*}$$

where

$$x_{ijk}^* = np_{ijk}^*, \text{ and } \hat{p} = x_{ijk}/n.$$

The distribution of $2nI(\hat{p}:p^*)$ is asymptotically chi-squared as shown in Kullback [1959], Kupperman [1957] with $(r-1)(c-1)(d-1)$ degrees of freedom. The same result is obtained if we note that p_{ijk}^* 's are RBAN estimators in the sense of Neyman [1949] of the cell probabilities under the hypothesis of no second-order interaction, or CAN estimators of Rao [1965, p. 288]. The degrees of freedom are calculated from the general principle of equivalence of degrees of freedom and the number of independent restraints imposed by the specified marginals. In this case the degrees of freedom are

$$(3.4) \quad rcd - 1 - (r-1)(c-1) - (r-1)(d-1) - (c-1)(d-1) - (r-1) - (c-1) - (d-1) \\ = (r-1)(c-1)(d-1).$$

For a four-way table, the relationship between various interactions corresponding to completely specified sets of marginals are given in table 3.1. The following notations are used in the computation of degrees of freedom:

$$(3.5) \quad N = rcdt - 1 = N_1 + N_2 + N_3 + N_4 \\ N_1 = (r-1) + (c-1) + (d-1) + (t-1) \\ N_2 = (r-1)(c-1) + (r-1)(d-1) + (r-1)(t-1) + (c-1)(d-1) + (c-1)(t-1) \\ + (d-1)(t-1) \\ N_3 = (r-1)(c-1)(d-1) + (r-1)(c-1)(t-1) + (r-1)(d-1)(t-1) \\ + (c-1)(d-1)(t-1) \\ N_4 = (r-1)(c-1)(d-1)(t-1).$$

TABLE 3.1. *Analysis of information—four-way table*

Information	For testing the hypothesis of	Degrees of freedom
$2nI(\hat{p} : \hat{p}_3^*)$ $2nI(\hat{p}_3^* : \hat{p}_2^*)$	No third-order interaction	$N - N_1 - N_2 - N_3 = N_4$ N_3
$2nI(\hat{p} : \hat{p}_2^*)$ $2nI(\hat{p}_2^* : \hat{p}_1^*)$	No second-order interaction	$N - N_1 - N_2$ N_2
$2nI(\hat{p} : \hat{p}_1^*)$ $2nI(\hat{p}_1^* : p_0)$	Independence (No first-order interaction)	$N - N_1$ N_1
$2nI(\hat{p} : p_0)$	Uniformity	N

In table 3.1 we have specified a complete set of marginals as fixed for each hypothesis. This restriction is clearly unnecessary. We shall define the p^* -distribution generated by a partial set

of marginals or a mixed set of marginals of different order as the no "mixed-order interaction" hypothesis. An example of mixed-order interaction is the conditional independence hypothesis given by (2.24), where only two of the two-way marginals are specified in a three-way table.

To simplify the notation for a step by step analysis of a four-way table, let us denote the quantity

$$-\sum_{ijkl} p_{ijkl}^* \ln p_{ijkl}^*$$

by $H(\quad)$ where the indices within the bracket stand for the marginals that are considered fixed. If all the two-way marginals are considered fixed, we may write H (given two-way), or other descriptive phrases with defined meanings. The symbol $\hat{I}(\quad)$ will be used to denote the difference between $H(\quad)$ and H (data).

We note that the " H " notation used here is the notation for entropy. For the case $\pi_{ijkl} = 1/rcdt$, the problem of minimizing $I(p:\pi)$ in (2.4) subject to certain restraints is equivalent to minimizing

$$(3.6) \quad \sum_{ijk} p_{ijkl} \ln p_{ijkl} + \ln rcdt,$$

or maximizing the entropy

$$(3.7) \quad -\sum_{ijkl} p_{ijkl} \ln p_{ijkl}$$

subject to the same restraints. The latter problem has been considered by Good [1963, 1965, 1966].

Since higher-order marginals determine all lower marginals, we have, corresponding to (2.18), the following:

$$(3.8) \quad \begin{aligned} H(\text{data}) &\leq H(\text{given 3-way}) = H_3 \\ &\leq H(\text{given 2-way}) = H_2 \\ &\leq H(\text{given 1-way}) = H_1 \\ &\leq H(\text{uniform}) = H_0 = \ln rcdt. \end{aligned}$$

Hence,

$$(3.9) \quad \begin{aligned} \hat{I}(\text{given 1-way}) &= H_1 - H(\text{data}) \\ &= [H_2 - H(\text{data})] + (H_1 - H_2) \\ &\geq \hat{I}(\text{given 2-way}) \end{aligned}$$

$$(3.10) \quad \begin{aligned} \hat{I}(\text{given 2-way}) &= H_2 - H(\text{data}) \\ &= [H_3 - H(\text{data})] + (H_2 - H_3) \\ &\geq \hat{I}(\text{given 3-way}) \geq 0. \end{aligned}$$

In table 3.2 the two-way marginals are added one by one to the four one-way marginals until all six are specified; then the three-way marginals are added one by one to the complete set of two-way marginals until all four are specified. The components of information are expressed in terms of the differences in entropies, and when possible, also in terms of the form that can be obtained through the convexity property of m.d.i.s.

TABLE 3.2. Interactions in a four-way table

Marginal restraints added	p_{ijkl}^*	Information	Form obtained by convexity property	d.f.
$p_{i\dots}, p_{\cdot j\dots}, p_{\cdot\cdot k\dots}, p_{\cdot\cdot\cdot l}$	$p_{i\dots}p_{\cdot j\dots}p_{\cdot\cdot k\dots}p_{\cdot\cdot\cdot l}$	$H(i, j, k, l) - H(\text{data})$	$\sum_{ijkl} \hat{p}_{ijkl} \ln \frac{\hat{p}_{ijkl}}{p_{i\dots}p_{\cdot j\dots}p_{\cdot\cdot k\dots}p_{\cdot\cdot\cdot l}}$	$rcdt - 1 - (r-1) - (c-1) - (d-1) - (t-1) = N - N_1$
$p_{\cdot\cdot kl}$	$p_{i\dots}p_{\cdot j\dots}p_{\cdot\cdot kl}$	$H(i, j, k, l) - H(i, j, kl)$ $= H(k, l) - H(kl)$ $H(i, j, kl) - H(\text{data})$	$\sum_{kl} \hat{p}_{\cdot\cdot kl} \ln \frac{\hat{p}_{\cdot\cdot kl}}{p_{\cdot\cdot k\dots}p_{\cdot\cdot\cdot l}}$ $\sum_{ijkl} \hat{p}_{ijkl} \ln \frac{\hat{p}_{ijkl}}{p_{i\dots}p_{\cdot j\dots}p_{\cdot\cdot kl}}$	$(d-1)(t-1)$ $N - N_1 - (d-1)(t-1)$
$p_{i\dots l}$	$p_{\cdot j\dots} \frac{p_{i\dots l} p_{\cdot\cdot kl}}{p_{\cdot\cdot\cdot l}}$	$H(i, j, kl) - H(j, kl, il)$ $= H(i, l) - H(il)$ $H(j, kl, il) - H(\text{data})$	$\sum_{il} \hat{p}_{i\dots l} \ln \frac{\hat{p}_{i\dots l}}{p_{i\dots}p_{\cdot\cdot\cdot l}}$ $\sum_{ijkl} \hat{p}_{ijkl} \ln \frac{\hat{p}_{ijkl}}{p_{\cdot j\dots} (p_{\cdot\cdot kl} p_{i\dots l} / p_{\cdot\cdot\cdot l})}$	$(r-1)(t-1)$ $N - N_1 - (d-1)(t-1) - (r-1)(t-1)$
$p_{ij\dots}$	$\frac{p_{ij\dots} p_{\cdot\cdot kl} p_{i\dots l}}{p_{i\dots} p_{\cdot\cdot\cdot l}}$	$H(j, kl, il) - H(kl, il, ij)$ $= H(i, j) - H(ij)$ $H(kl, il, ij) - H(\text{data})$	$\sum_{ij} \hat{p}_{ij\dots} \ln \frac{\hat{p}_{ij\dots}}{p_{i\dots} p_{\cdot j\dots}}$ $\sum_{ijkl} \hat{p}_{ijkl} \ln \frac{\hat{p}_{ijkl}}{\frac{p_{ij\dots} p_{\cdot\cdot kl} p_{i\dots l}}{p_{i\dots} p_{\cdot\cdot\cdot l}}}$	$(r-1)(c-1)$ $N - N_1 - (d-1)(t-1) - (r-1)(t-1) - (r-1)(c-1)$
$p_{\cdot jk\dots}$		$H(kl, il, ij) - H(kl, il, ij, jk)$ $H(kl, il, ij, jk) - H(\text{data})$	$\sum_{jk} \hat{p}_{\cdot jk\dots} \ln \frac{\hat{p}_{\cdot jk\dots}}{\sum_{il} \frac{p_{ij\dots} p_{\cdot\cdot kl} p_{i\dots l}}{p_{i\dots} p_{\cdot\cdot\cdot l}}}$	$(c-1)(d-1)$ $N - N_1 - (d-1)(t-1) - (r-1)(t-1) - (r-1)(c-1) - (c-1)(d-1)$

TABLE 3.2. *Interactions in a four-way table* — Continued

Marginal restraints added	p_{ijkl}^*	Information	Form obtained by convexity property	d.f.
$p_{i \cdot k \cdot}$		$H(kl, il, ij, jk) - H(kl, il, ij, jk, ik)$ $H(kl, il, ij, jk, ik) - H(\text{data})$	Iteration on indicated two-way marginals	$(r-1)(d-1)$ $N - N_1 - (d-1)(t-1)$ $- (r-1)(t-1)$ $- (r-1)(c-1)$ $- (c-1)(t-1)$ $- (r-1)(d-1)$
$p_{j \cdot l}$	Second-order interaction	$H(kl, il, ij, jk, ik) - H(\text{given two-way})$ $H(kl, il, ij, jk, ik, jl) - H(\text{data})$		$(c-1)(t-1)$ $N - N_1 - N_2$
Six two-way marginals		$H(\text{given two-way}) - H(\text{data})$		$N - N_1 - N_2$
$p_{ijk \cdot}$		$H(\text{given two-way}) - H(il, jl, kl, ijk)$ $H(il, jl, kl, ijk) - H(\text{data})$		$(r-1)(c-1)(d-1)$ difference
$p_{ij \cdot l}$		$H(il, jl, kl, ijk) - H(ijk, ijl)$ $H(ijk, ijl) - H(\text{data})$		$(r-1)(c-1)(t-1)$ difference
$p_{i \cdot k l}$		$H(ijk, ijl,) - H(ijk, ijl, ikl)$ $H(ijk, ijl, ikl) - H(\text{data})$		$(r-1)(d-1)(t-1)$ difference
$p_{\cdot jkl}$	Third-order interaction	$H(ijk, ijl, ikl) - H(\text{given three-way})$ $H(\text{given three-way}) - H(\text{data})$		$(c-1)(d-1)(t-1)$ $N - N_1 - N_2 - N_3$ $= (r-1)(c-1)(d-1)(t-1)$ $= N_4$

We note in table 3.2 that the addition of each two-way marginal restraint generates a hypothesis of two-way independence of the corresponding marginal table, or the conditional two-way independence given one or more marginals. Beginning with the fourth marginal restraint, however, these hypotheses can no longer be expressed in an explicit closed form and the respective p^* -distributions will have to be generated by the iteration procedure.

3.2. Some Typical Second-Order No-Interaction Hypotheses

In many practical applications, the hypothesis of interest is usually suggested by the physical relationship involved in the problem, and the result of the test admits natural interpretations. The no second-order interaction in a three-way table originated in this manner, i.e., it is a test of the sameness of the measure of association between R and C classifications over categories of D [Simpson 1951].

With the addition of another dimension in a four-way table, there are a number of mixed second-order interactions to which there are no corresponding ones in a three-way table. Some of the typical ones are described below.

The interaction of a one-way by two-way interaction over the fourth classification is a mixed second-order interaction. There are six such second-order interactions, corresponding to the six three-way tables with different marginal probabilities, and possibly different dimensions, that can be constructed from a four-way table. Since symmetry is a property of our procedure, these interactions could also be considered as the interaction of one-way by one-way interaction over categories of the two remaining classifications. For example, the second-order interaction $(DT \times R)(C)$ is the same as the second-order interaction $(R \times C)(DT)$.

The analysis of information table for these second-order interactions is given in table 3.3, using the convexity property of the m.d.i.s. to indicate:

- (1) how this second-order interaction can be derived, and
- (2) the particular marginals which must be specified for the iteration procedure, viz., $p_{ij..}$, $p_{.jkl}$, and $p_{i..kl}$ in this case.

We note that if we consider

$$p_{ijkl}^{(0)} = \frac{p_{ij..} \cdot p_{.jkl}}{p_{.j..}},$$

then

$$\begin{aligned} p_{ijkl}^{(1)} &= \frac{p_{i..kl}}{p_{i..kl}^{(0)}} p_{ijkl}^{(0)} \\ &= \frac{p_{ij..} \cdot p_{.jkl}}{p_{.j..}} \cdot \frac{p_{i..kl}}{\sum_j \frac{p_{ij..} \cdot p_{.jkl}}{p_{.j..}}}, \end{aligned}$$

which is exactly the denominator appearing in the expression for second-order interaction. Hence the convexity property of the m.d.i.s. is useful in giving an explicit expression for the p_{ijkl}^* value after the first iteration. This agreement is not surprising since we are utilizing two distinct but consistent properties of the m.d.i.s.

Viewing the four-way table in another perspective, there are four distinct second-order interactions defined as the interaction of the three-way interaction over the fourth classification. The analysis for $(R \times D \times T)(C)$ is shown in table 3.4. The marginals to be considered as specified are

$$p_{ij..}, p_{.jk..}, p_{.j..t}, \text{ and } p_{i..kl}.$$

TABLE 3.3. *Mixed-order interactions in a four-way table*

Component due to	Information	d.f.
$(DT \times R)(C)$ Second-order interaction ^a	$2\sum_{ijkl} p_{ijkl} \ln \frac{p_{ijkl}}{\frac{p_{ij..} \cdot p_{.jkl}}{p_{.j..}} \cdot \frac{p_{i..kl}}{\sum_j \frac{p_{ij..} \cdot p_{.jkl}}{p_{.j..}}}}$	$(c-1)(r-1)(dt-1)$
$(DT, R)(C)$ Three-way marginals interaction with column	$2\sum_{ikt} p_{i..kl} \ln \frac{p_{i..kl}}{\sum_j \frac{p_{ij..} \cdot p_{.jkl}}{p_{.j..}}}$	$(r-1)(dt-1)$
$DT \times R C$ Two-way by one-way interaction, given column	$2\sum_{ijk} p_{ijkl} \ln \frac{p_{ijkl}}{p_{ij..} \cdot p_{.jkl} / p_{.j..}}$	$c(r-1)(dt-1)$

^a Form shown is the second-order interaction after first iteration.

Conceptually a third-order interaction in a four-way table may be defined as the interaction of the second-order interactions of the three classifications over categories of the fourth classification. The analysis following this line of thought is given in table 3.5, showing the marginals to be specified are

$$p_{ijk..}, p_{i..kl}, p_{ij..l}, \text{ and } p_{.jkl}.$$

It is clear that we do not have a direct counterpart of third-order interaction in the classical hypotheses. The interpretation of no third-order interaction also becomes obscure in the conventional sense. We propose, therefore, to consider a hypothesis represented by the p^* -distribution as that of a generalized independence (generalized no interaction, no association) among the classifications with given fixed marginals, and give a unified interpretation in the following subsection.

TABLE 3.4. *Mixed-order interactions in a four-way table*

Component due to	Information	d.f.
$(R \times D \times T)(C)$ Second-order interaction ^a	$2\sum_{ijkl} p_{ijkl} \ln \frac{p_{ijkl}}{\frac{p_{ij..} \cdot p_{.jk} \cdot p_{.j..l}}{(p_{.j..})^2} \cdot \frac{p_{i..kl}}{\sum_j \frac{p_{ij..} \cdot p_{.jk} \cdot p_{.j..l}}{(p_{.j..})^2}}}$	$(c-1)(rdt - r - d - t + 2)$
$(R, D, T)(C)$ Three-way marginals interaction with column	$2\sum_{ikt} p_{i..kl} \ln \frac{p_{i..kl}}{\sum_j \frac{p_{ij..} \cdot p_{.jk} \cdot p_{.j..l}}{(p_{.j..})^2}}$	$rdt - r - d - t + 2$
$R \times D \times T C$ Three-way interaction, given column	$2\sum_{ijk} p_{ijkl} \ln \frac{p_{ijkl}}{\frac{p_{ij..} \cdot p_{.jk} \cdot p_{.j..l}}{(p_{.j..})^2}}$	$c(rdt - r - d - t + 2)$

^a Form shown is the second-order interaction after the first iteration.

TABLE 3.5. *Third-order interaction derived from the convexity property*

Component due to	Information	d.f.
$(R \times T)(D)(C)$ Third-order interaction	$2 \sum_{ijkl} p_{ijkl} \ln \frac{p_{ijkl}}{\frac{p_{i \cdot kl}}{\sum_j \frac{p_{ijk} \cdot p_{\cdot jkl}}{p_{\cdot jk}} \cdot \frac{p_{ij \cdot l}}{\sum_k \frac{p_{ijk} \cdot p_{\cdot jkl}}{p_{\cdot jk}}}} \cdot \frac{p_{ij \cdot l}}{p_{\cdot jk}} \cdot \frac{p_{ijk} \cdot p_{\cdot jkl}}{p_{\cdot jk}}}$	$(r-1)(c-1)(d-1)(t-1)$
$(R, D, T)(C)$ Three-way marginals interaction with column	$2 \sum_{ikl} p_{i \cdot kl} \ln \frac{p_{i \cdot kl}}{\sum_j \frac{p_{ijk} \cdot p_{\cdot jkl}}{p_{\cdot jk}} \cdot \frac{p_{ij \cdot l}}{\sum_k \frac{p_{ijk} \cdot p_{\cdot jkl}}{p_{\cdot jk}}}}$	$(r-1)(d-1)(t-1)$
$(R \times T)(D) C$ Second-order interaction, given column ^a	$2 \sum_{ijk} p_{ijk} \ln \frac{p_{ijk}}{\frac{p_{ijk} \cdot p_{\cdot jkl}}{p_{\cdot jk}} \cdot \frac{p_{ij \cdot l}}{\sum_k \frac{p_{ijk} \cdot p_{\cdot jkl}}{p_{\cdot jk}}}}$	$c(r-1)(d-1)(t-1)$
$(R, T)(D) C$ Two-way marginals interaction with depth, given column	$2 \sum_{ijl} p_{ij \cdot l} \ln \frac{p_{ij \cdot l}}{\sum_k \frac{p_{ijk} \cdot p_{\cdot jkl}}{p_{\cdot jk}}}$	$c(r-1)(t-1)$
$R \times T CD$ One-way by one-way interaction, given CD	$2 \sum_{ijk} p_{ijk} \ln \frac{p_{ijk}}{p_{ijk} \cdot p_{\cdot jkl} / p_{\cdot jk}}$	$cd(r-1)(t-1)$

^a Form shown is the second-order interaction after first iteration.

3.3. Logarithmic Additivity and a General Interpretation of the No-Interaction Hypothesis

In theorem 2.1A, we give the p^* -distribution for no second-order interaction as

$$(3.11) \quad \ln \frac{p_{ijk}^*}{\pi_{ijk}} = \ln a_{ij} + \ln b_{jk} + \ln c_{ik}$$

where a_{ij} , b_{jk} and c_{ik} are functions of the three two-way marginals. In this form, the logarithms of the cell probabilities representing the no second-order interaction hypothesis are seen to be the sum of a constant and the logarithms of contributions from each of the specified marginals. Similarly, for the test of the hypothesis of uniform distribution, we have

$$(3.12) \quad \ln p_{ijk}^* = \ln \frac{1}{rcd},$$

and for the test of the hypothesis of three-way independence,

$$(3.13) \quad \ln p_{ijk}^* = \ln \frac{1}{rcd} + \ln a_i + \ln b_j + \ln c_k.$$

Hence, if we consider the difference in the logarithms of the cell probabilities between the independence hypothesis and the uniform distribution hypothesis as represented by a row effect, a

column effect, and a depth effect, then the striking similarities between our approach to contingency table analysis and the approach used in analysis of variance become immediately obvious.

To begin with, both analyses deal with multifactor multiresponse data. In the linear hypothesis, an additive model is assumed; in the analysis of contingency tables, a logarithmic additive model is assumed. In both cases the appropriate test statistics can be obtained by minimizing the discrimination information [ch. 10, 11, Kullback 1959]. The residuals in the usual linear analysis represent the difference between the observed values and the values computed from the model using the estimated values of the parameters; the residuals in our analysis represent the differences between the observed cell frequencies and the cell frequencies x^* computed under each particular hypothesis. The main effects and the various interactions in the analysis of variance also find corresponding counterparts in our first-order, second-order, and higher-order interactions. Darroch [1961], Bhakpar [1961], Lindley [1964], and Mantel [1966] have all suggested some analogy between the two types of analyses with a view to simplifying the analysis of multidimensional contingency tables. We remark that the main difference between the two types of analysis is that the marginal restraints requirements in the contingency table analysis which necessitate the iteration procedure are not present in the analysis of variance.

We may consider the complete sample table to contain all the "information" available from the particular experiment. In the process of analysis, we aim to express the sample table in a reduced number of parameters represented by the marginal totals as expressed in (3.11) to (3.13). In other words, we are interested in knowing how much of this total information is contained in a summary consisting of sets of marginal tables.

If there is no first-order interaction, i.e., independence of all classifications, then all the information is contained in the first-order marginals in the sense that given these marginals, the complete table can be constructed to within sampling error. If the first-order interaction is significant, but there is no second-order interaction, then the set of two-way marginals will be required to summarize the data adequately. The use of two-way tables to summarize multiway classification data is a rather common practice, and the implied assumption is therefore "no second- and higher-order interactions."

A direct consequence of this interpretation is that the analysis can be reduced to that of the set of marginal tables if there is no interaction of the same order.

We remark that the set of marginal tables must be considered jointly for proper interpretation, and if one or more of these tables show significant interactions, the results of tests of the remaining tables could lead to erroneous conclusions. An example of such a case was given in Simpson [1951].

The above interpretation is not restricted to complete sets of marginals. If the p^* -distribution computed from three out of the six two-way marginals in a four-way table is found to be "close enough" to the \hat{p} -distribution by our test, the three two-way marginal tables could be considered as containing essentially all the information in the four-way table. The analysis can therefore be performed on these marginal tables and the complexity of the problem reduced. For example, the analysis given in table 3.3 for a four-way table may be reduced to that of one two-way and two three-way tables, and that in table 3.4 to that of three two-way tables and one three-way table, provided that the corresponding interactions are found to be of no significance.

A useful by-product obtained as a result of our computer routine is that the set of residuals, $x - x^*$, are computed for each interaction hypothesis. Inspection and analysis of these residuals may be used as an aid "in assessing the validity or appropriateness of the conventional analysis" as recommended by Anscombe and Tukey [1963], in view of the indefiniteness and complexity of objectives of statistical analysis of multiresponse data.

The analysis of categorical data may well follow this general philosophy and take advantage of some of the developed techniques for the analysis of residuals. In fact, in a goodness-of-fit test, if the computed X^2 shows significance, we usually look at the larger discrepancies between the observed and expected values of the cell frequencies and seek for an explanation. However, this practice has been restricted mainly to one-way tables. A plausible reason for the lack of such study in higher-order tables could be that the computation of expected frequencies becomes complicated

and unmanageable. Consequently, the analysis is considered complete with a formal test of significance.

With the iterative procedure now available for computing the expected frequencies for each particular hypothesis, we could examine these residuals for a number of interesting features, such as:

- (a) outliers, or errors in counting and recording,
- (b) the physical interpretation of departure from the particular hypothesis,
- (c) trend over categories in a classification, particularly if these categories are arranged in a natural sequence of order of magnitude [Cochran 1954, pp. 434-436], and
- (d) the agreement of these residuals, squared and weighted by the inverses of the expected frequencies, with the corresponding X^2 distribution.

We have therefore computed and printed these residuals, $x - x^*$, or normalized residuals in table forms as a by product of our computation procedures to facilitate visual examination.

4. Computer Programs and Examples

4.1. Computation and Iteration Programs

The iterative computation process described above for the calculation of cell probabilities, or cell frequencies, representing that of no interaction when certain marginals are considered fixed, is ideally suited for electronic computer operation. A program in Fortran V has been prepared for this purpose.⁴ A brief description of this program is given below.

(1) The program is written in double precision mode for the computation of quantities of the form $2\sum x \ln x$. These quantities are useful in testing certain hypotheses as illustrated in Kullback, Kupperman, and Ku [1962a, 1962b]. The quantity $0 \ln 0$ is defined as zero.

(2) Input cards are provided for the specification of:

(a) dimension M of the table, and number of categories within each dimension, with $2 \leq M \leq 4$, and $r \times c \times d \times t \leq 10^4$.

(b) maximum number of complete cycles of iterative computation, and the agreement desired between the original given marginals and the computed marginals. Tentatively the maximum number of cycles is set at 20 and the agreement optionally at 0.100, 0.010, and 0.001.

(c) the choice of the set of marginals if these marginals are not a complete set of one-, two-, or three-way marginals. Iterative computation for the complete sets of marginals is automatically performed.

(3) The data cards for the table are read in by column within each row, row \times column within each depth, and row \times column \times depth within each level. Title cards for each of the classifications are provided.

(4) The following notations are used in the output

X(IJKL) original data

Y(IJKL) cell frequencies corresponding to no first-order interaction.

Z(IJKL) cell frequencies corresponding to no second-order interaction.

W(IJKL) cell frequencies corresponding to no third-order interaction.

U1(IJKL)
U2(IJKL)} cell frequencies corresponding to specified marginals.

(5) Outputs of the program for a four-way table are in the order listed below. For two- and three-way tables, the input cards will adjust the outputs accordingly.

(a) Titles of classifications.

(b) Original table X(IJKL) in the form of two-way tables.

(c) All marginal three-way, two-way, one-way tables and the grand total.

(d) All 16 sums of quantities of the form $2\sum X \ln X$.

⁴ We are indebted to Mrs. Ruth Varner, Statistical Engineering Laboratory, National Bureau of Standards, for the preparation of this program.

(e) Number of complete cycles of iterations performed and the decimal agreement between marginals.

(f) Tables of normalized residuals = $\{X(IJKL) - Y(IJKL)\} / \sqrt{Y(IJKL)}$

$$2\sum_Y \ln Y$$

$$\text{first-order interaction} = 2\sum X \ln X - 2\sum Y \ln Y$$

$$\text{chi-squared} = \sum \frac{(X - Y)^2}{Y}$$

(g) Print-outs under (f) are repeated for Z, and W, and for U1, U2 when specified.

Samples of the output are shown in the appendix for the four-way contingency tables used in examples 1 and 2.

4.2. Examples

In the literature there are a number of "classical" examples which have been used to demonstrate tests of no second-order interaction in three-way tables. These examples are collected and listed in table 4.1 where the values of the m.d.i.s. for no second-order interaction are compared with results obtained by other investigators. A number of interesting features are noted.

(1) The maximum number of complete cycles of iteration used was 10 for the $2 \times 2 \times 12$ table due to Snedecor. For the others, 6 to 7 cycles are sufficient for agreement of specified marginals to the third decimal place.

(2) The values of $2\hat{I}$ for no second-order interaction agrees very well with the values of X^2 computed through the solutions of systems of simultaneous equations of third degree, i.e., solutions with 2-way marginals considered as fixed. Solutions based on unrestricted maximum likelihood estimates are, however, somewhat lower than our values.

(3) None of the second-order interactions computed reached the 5-percent level of significance. By the interpretation given in subsection 3.3, conclusions drawn from analysis of the three 2-way marginal tables are valid for the 3-way table.

EXAMPLE 1. Ries and Smith [1963] reported an experiment comparing two detergents, a new product X and a standard product M. The three classifications were water softness, at three levels, temperature, at two levels, and a factor corresponding to previous experience and no previous experience with detergent M. This is a $2 \times 2 \times 2 \times 3$ experiment with

R: preference	$i=1$	X
	2	M
C: water temperature	$j=1$	low
	2	high
D: previous use	$k=1$	nonuser
	2	user
T: water softness	$l=1$	hard
	2	medium
	3	soft

Ries and Smith used a series of chi-squared tests in their paper; Cox and Lauh [1967] reexamined the data recently employing a graphical approach. The data and computations are shown in the print-out sample A in the appendix.

TABLE 4.1. *Comparison of results—second-order interaction in three-way tables*

Example	Value of $2\hat{I}$ with marginal agreement to			Other results		
	0.10	0.01	0.001			
Bartlett [1935] $2 \times 2 \times 2$	2.357(4) ^a	2.296(6)	2.294(7)	2.27 2.298 1.93 2.26 2.27 8.89 0.083 7.603	X^2 $2\hat{I}$ X^2 Y^2 Z^2 X_m^2 X_1^2 $2\hat{I}$	Bartlett. Computed from Bartlett's solution. Goodman [1964b]. Goodman [1964b]. Goodman [1964b]. Koch [1968]. Koch [1968] different model. Kullback [1959, without iteration].
Kastenbaum and Lamphiear [1959] $5 \times 3 \times 2$ $3 \times 2 \times 5$ $2 \times 3 \times 5$	3.1838(2) 3.1660(4) 3.1609(3)	3.1592(4) 3.1600(5) 3.1591(4)	3.1588(7)	3.158 3.640 3.13 2.8 3.12	X^2 $2\hat{I}$ Y^2 \hat{Y}^2 X^2	Kastenbaum [1959] Darroch [1962]. KKK [1962b, without iteration]. Goodman [1964b]. Goodman [1964b]. Plackett [1962].
Snedecor, as quoted on p. 184, Kullback [1959] $2 \times 2 \times 12$		7.7157(10)		12.608 15.492 7.59 7.45 7.37 5.18	$2\hat{I}$ $2\hat{I}$ X^2 X^2 Y^2 Z^2	Kullback (without iteration). Kullback (without iteration). Norton [1945]. Goodman [1964b]. Goodman [1964b]. Goodman [1964b].
Kullback [1959], prob. 13.10, p. 188 $2 \times 2 \times 2$		0.007(5)		0		Theoretical.
Kullback [1959], table 12.2, p. 180 $2 \times 4 \times 2$		7.584(2)		7.570	$2\hat{I}$	Kullback (algebraic).
Kilkberg, Narragon and Campbell [1964] Bhaskar and Koch $2 \times 2 \times 2$			0.0704(5) 0.0430(6) 3.5710(6)	0.071 0.0435 3.3917	X_f^2 X_e^2 X_e^2	Koch [1968]. Koch [1968]. Koch [1968].
Schotz [1966] Bhaskar and Koch $2 \times 2 \times 4$			10.511(7)	7.22	X_G^2	Koch [1968, using index of order association].

^a Number of complete cycles of iteration.

The analysis of information table corresponding to table 3.1 is shown in table 4.2 below.

TABLE 4.2

Components of information	Information	d.f.
Third-order interaction $2nI(p_3^*:p_2^*)$	0.739 9.108	2 7
Second-order interaction $2nI(p_2^*:p_1^*)$	9.847 33.081	9 9
Four-way independence	42.928	18

Neither the third-order nor the second-order interactions reached significance at $\alpha=0.10$. Hence we conclude that the analysis of the six two-way tables will yield the desired information. The numerical values of these six interactions are computed in two ways for comparison in table 4.3. The first set is computed directly from the six two-way tables. The second set is computed by using the analysis $2nI(p_2^*:p_1^*)=2\sum_{ij}X_{ij}\ln a_{ij} + \dots + 2\sum_{kl}X_{kl}\ln f_{kl}$. The sum of the first set, 33.763, should equal the component $2nI(p_2^*:p_1^*)$ in table 4.2. The difference between the two sums represents the effect of the marginal restraints.

TABLE 4.3

Components of information	Information		d.f.
Preference and water temp. ($R \times C$).....	4.361	4.393	1
Preference and previous use ($R \times D$).....	20.581	19.920	1
Preference and water softness ($R \times T$).....	.395	.424	2
Water temp. and previous use ($C \times D$).....	1.252	1.314	1
Water temp. and water softness ($C \times T$).....	6.099	6.089	2
Previous use and water softness ($D \times T$).....	1.075	.943	2
	33.763	33.083	9

The main conclusion here is that preference is highly dependent on previous use, and to a certain extent dependent on the water temperature. The water temperature effect depends somewhat on degree of softness of water. The nonsignificance of ($C \times D$) and ($D \times T$) shows that the samples of previous user and nonprevious user of M are not biased with respect to water temperature and water softness.

We also include here analysis of information tables 4.4 and 4.5 corresponding to tables 3.3 and 3.4 respectively.

TABLE 4.4

Components	Information	d.f.
$(CT \times R)(D)$	8.059	5
$(CT, R)(D)$	4.185	5
$CT \times R D$	12.244	10

TABLE 4.5

Components	Information	d.f.
$(R \times C \times T)(D)$	9.725	7
$(R, C, T)(D)$	10.294	7
$R \times C \times T D$	20.019	14

Since none of these components of information reached significance at $\alpha = 0.10$, we conclude that the interaction between preference and water temperature-water softness are not different for the previous user and nonuser groups. If this conclusion is accepted, then separate analyses of previous user group and nonuser group appear to be unnecessary.

EXAMPLE 2.

For the second sample we use the survey results as reported by Hoyt, Krishnaiah, and Torrance [1959], analyzed also in Kullback, Kupperman and Ku [1962a] for nine hypotheses of independence and conditional independence. The four classifications are:

	Categories
D: high school ranks	3
C: post high school status	4
T: sex	2
R: father's occupational level	7

The data (sample B, appendix) showed considerable heterogeneity and all the nine hypotheses tested in the above analysis gave highly significant results. We continue the analysis in table 4.6. All the interactions are again highly significant excepting the third-order interaction for which $p \doteq 0.15$.

Since the second-order interaction is highly significant, we analyze the difference between second-order and third-order interaction into its component parts in table 4.7, the second set in accordance with the second part of table 3.2. These four components represent second-order interactions in the four three-way tables subject to the three-way marginal restraints. The first set corresponds to the analysis $2nI(p^*:p^*) = 2\sum_{ijk} X_{ijk} \ln a_{ijk} + \dots + 2\sum_{jkl} X_{jkl} \ln d_{jkl}$.

We note that all the second-order interactions are significant when the R classification is involved, i.e., the interactions $C \times D$, $C \times T$, and $D \times T$ are different for different occupational levels of fathers' occupations. These results, and the fact that there is an unusually larger number of girls than boys for the third level of fathers' occupation as shown in table for $X(I^{**}L)$, suggest that the counts for this level may be suspect.

TABLE 4.6

Components of information	Information	d.f.
$2nI(\hat{p} : p_3^*)$	44.793	36
$2nI(p_3^* : p_2^*)$	127.464	72
$2nI(\hat{p} : p_2^*)$	172.257	108
$2nI(p_2^* : p_1^*)$	3320.858	47
$2nI(\hat{p} : p_1^*)$	3493.115	155

TABLE 4.7

Marginal restraints added	Information		d.f.
All two-way marginals	172.257	172.257	108
RCD	53.841	52.267	36
		119.990	72
RCT	45.161	44.630	18
		75.360	54
RDT	25.477	27.588	12
		47.772	42
CDT	2.985	2.979	6
All three-way marginals	44.793	44.793	36

TABLE 4.8

All two-way marginals		109.521	91
RCD		29.521	30
		80.000	61
RCT		27.738	15
		52.262	46
RDT		11.986	10
		40.276	36
CDT		.673	6
All three-way marginals		39.603	30

Table 4.8 is an analysis of the data with the third level of fathers' occupation deleted. The second-order interaction is still significant at $\alpha=0.10$, but not at $\alpha=0.05$. Of the components of second-order interaction, however, only RCT remains significant. The interpretation that the interactions between post high school status and sex are different for different levels of fathers' occupational level appears to be a reasonable one.

We would suggest, therefore, that the data for the third level of fathers' occupation be rechecked.

5. Summary

Using the basic notions of information theory, we have developed in the above sections a unified approach to the analysis of multiway contingency tables. Under this approach the principle of minimum discrimination information is proposed and used to generate hypotheses of interest. It is shown that all classical hypotheses for contingency tables can be generated through the use of this principle when certain marginals are considered fixed.

For each set of fixed marginals, a unique set of cell probabilities $\{p^*\}$ is generated by minimizing the discrimination information. The set of $\{p^*\}$ corresponds to the cell probabilities representing no-interaction, and typically can be expressed in a logarithmic linear form:

$$\ln p_{ijk}^* = \text{const.} + \ln a_{ij} + \ln b_{jk} + \ln c_{ik},$$

where a_{ij} , b_{jk} , and c_{ik} are functions of cell probabilities of the corresponding fixed two-way marginal tables. The difference between the set of cell probabilities estimated from data and $\{p^*\}$ is therefore a measure of interaction.

If the complete set of one-way marginals are considered fixed, the set of $\{p_i^*\}$ represents cell probabilities under the independence hypothesis. If the complete set of two-way marginals are considered fixed, the set of $\{p_{ij}^*\}$ are the cell probabilities representing no second-order interaction. In this sense the higher-order no-interaction hypotheses can be considered as hypotheses of "generalized" independence, a concept which unifies the many attempts in the formulation of second-order interaction described in brief in the introductory section.

The relationship between minimum discrimination information and maximum entropy is examined and the analogy between the proposed analysis and the analysis of variance using least squares theory is noted. An interpretation of the no-interaction hypothesis as equivalent to that "the given marginal tables are sufficient and contain all the information of the full table" reduces the dimension of the table, and hence also the complexity of the analysis.

The expression for p^* for given marginals is given in theorem 2.1 and the convergence of the iterative computation procedure to the unique set of $\{p^*\}$ in theorem 2.2.

Analysis of information tables for four-way tables are given for first-, second-, and third-order interactions, and also for selected mixed-order interactions. A Fortran program to aid in the computation has been prepared.

Two illustrative examples in the analysis of four-way tables are included.

6. Appendix. Samples of Selected portions of Computer Print-Out

SAMPLE A EXAMPLE 1 (Tables of residuals suppressed)

R	PREFERENCE M OR X	I=1, 2
C	WATER TEMPERATURE	J=1, 2
D	PREVIOUS USER OR NONUSER OF M	K=1, 2
T	WATER SOFTNESS	L=1, 3

ORIGINAL TABLES

X(IJ11)
68 42
42 30

X(IJ12)
66 33
50 23

X(IJ13)
63 29
53 27

X(IJ21)
37 24
52 43

X(IJ22)
47 23
55 47

X(IJ23)
57 19
49 29

MARGINAL TABLES

THREE-WAY TABLES

X(IJ1*)		X(*J1L)		X(1J*L)			X(I*1L)			
197	104	110	116	116	105	113	120	110	99	92
145	80	72	56	56	66	56	48	72	73	80

X(IJ2*)		X(*J2L)		X(2J*L)			X(I*2L)			
141	66	89	102	106	94	105	102	61	70	76
156	119	67	70	48	73	70	56	95	102	78

TWO-WAY TABLES

X(IJ**)		X(I*K*)		X(I**L)		
338	170	301	207	171	169	168
301	199	225	275	167	175	158

X(*JK*)		X(*J*L)			X(**KL)		
342	297	199	218	222	182	172	172
184	185	139	126	104	156	172	154

ONE-WAY TABLES

X(I***)		X(*J**)		X(**K*)		X(***L)		
508	500	639	369	526	482	338	344	326

TOTAL

X(****)
1008

PRINT OF SUMS

SUM 2X(IJKL)LNX(IJKL)= .7653768886020237 + 004

SUM 2X(IJK*)LNX(IJK*)= .9853563803650491 + 004

SUM 2X(*JKL)LNX(*JKL)= .9018264486617197 + 004

SUM 2X(IJ*L)LNX(IJ*L)= .9017229799353486 + 004

SUM 2X(I*KL)LNX(I*KL)= .8962284823389048 + 004

SUM 2X(IJ**)LNX(IJ**)=.1122496619356278+005
 SUM 2X(I*K*)LNX(I*K*)=.1116989524449928+005
 SUM 2X(I**L)LNX(I**L)=.1033087165370435+005
 SUM 2X(*JK*)LNX(*JK*)=.1122371544867871+005
 SUM 2X(*J*L)LNX(*J*L)=.1040972402615119+005
 SUM 2X(**KL)LNX(**KL)=.1033340920997269+005
 SUM 2X(I***)LNX(I***)=.1254477724916192+005
 SUM 2X(*J**)LNX(*J**)=.1261792581599639+005
 SUM 2X(**K*)LNX(**K*)=.1254663500174484+005
 SUM 2X(***L)LNX(***L)=.1172779757827051+005

2N LN N=.1394209847244072+005

2Y LN Y=.7610840227851502+004
 FIRST-ORDER INTERACTION=.4292865816+002
 CHI-SQUARED=.4390224840+002

2Z LN Z=.7643922243020641+004
 SECOND-ORDER INTERACTION=.9846642999+001
 CHI-SQUARED=.9870614978+001

2W LN W=.7653029670355177+004
 THIRD-ORDER INTERACTION=.7392156650+000
 CHI-SQUARED=.7379092751+000

SPECIFIED MARGINALS I*K* *JKL
 NO. OF ITERATIONS=1 CYCLE
 AGREEMENT BETWEEN MARGINALS TO .100-01

2(U1) LN (U1)=.7641524729371637+004
 INTERACTIONS (U1)=.1224415664+002
 CHI-SQUARED=.1220141783+002

SPECIFIED MARGINALS I*K* *JKL IJ*L
 NO. OF ITERATIONS=3 CYCLES
 AGREEMENT BETWEEN MARGINALS TO .100-01

2(U2) LN (U2)=.7645709639307871+004
 INTERACTIONS (U2)=.8059246712+001
 CHI-SQUARED=.8054468429+001

SPECIFIED MARGINALS I*K* *JK* **KL
 NO. OF ITERATIONS=1 CYCLE
 AGREEMENT BETWEEN MARGINALS TO .100-01

2(U3) LN (U3)= .7633749899660993 + 004
 INTERACTIONS (U3)= .2001898635 + 002
 CHI-SQUARED= .2069489400 + 002

SPECIFIED MARGINALS I*K* *JK* **KL IJ*L
 NO. OF ITERATIONS= 3 CYCLES
 AGREEMENT BETWEEN MARGINALS TO .100-01
 2(U4) LN (U4)= .7644043799921927 + 004
 INTERACTIONS (U4)= .9725086098 + 001
 CHI-SQUARED= .9720352364 + 001

SAMPLE B EXAMPLE 2 (Mixed order interactions not shown)

R FATHER OCCUPATIONAL LEVEL I= 1, 2, 3, 4, 5, 6, 7
 C POST HIGH SCHOOL STATUS J= 1, 2, 3, 4
 D HIGH SCHOOL RANKS K= 1, 2, 3
 T SEX L= 1, 2

ORIGINAL TABLES

	X(IJ11)				X(IJ21)				X(IJ31)			
87	3	17	105	216	4	14	118	256	2	10	53	
72	6	18	209	159	14	28	227	176	8	22	95	
52	17	14	541	119	13	44	578	119	10	33	257	
88	9	14	328	158	15	36	304	144	12	20	115	
32	1	12	124	43	5	7	119	42	2	7	56	
14	2	5	148	24	6	15	131	24	2	4	61	
20	3	4	109	41	5	13	88	32	2	4	41	

	X(IJ12)				X(IJ22)				X(IJ32)			
53	7	13	76	163	30	28	118	309	17	38	89	
36	16	11	111	116	41	53	214	225	49	68	210	
52	28	49	521	162	64	129	708	243	79	184	448	
48	18	29	191	130	47	62	305	237	57	63	219	
12	5	10	101	35	11	37	152	72	20	21	95	
9	1	15	130	19	13	22	174	42	10	19	105	
3	1	6	88	25	9	15	158	36	14	19	93	

MARGINAL TABLES

THREE-WAY TABLES

	X(IJ1*)				X(IJ2*)				X(IJ3*)			
140	10	30	181	379	34	42	236	565	19	48	142	
108	22	29	320	275	55	81	441	401	57	90	305	
104	45	63	1062	281	77	173	1286	362	89	217	705	
136	27	43	519	288	62	98	609	381	69	83	334	
44	6	22	225	78	16	44	271	114	22	28	151	
23	3	20	278	43	19	37	305	66	12	23	166	
23	4	10	197	66	14	28	246	68	16	23	134	

THREE-WAY TABLES – Continued

X(*J1L)		X(*J2L)		X(*J3L)	
365	213	760	650	793	1164
41	76	62	215	38	246
84	133	157	346	100	412
1564	1218	1565	1829	678	1259

X(1J*L)		X(2J*L)		X(3J*L)	
559	525	407	377	290	457
9	54	28	106	40	171
41	79	68	132	91	362
276	283	531	535	1376	1677

X(4J*L)		X(5J*L)		X(6J*L)		X(7J*L)	
390	415	117	119	62	70	93	64
36	122	8	36	10	24	10	24
70	154	26	68	24	56	21	40
747	715	299	348	340	409	238	339

X(I*1L)		X(I*2L)		X(I*3L)	
212	149	352	339	321	453
305	174	428	424	301	552
624	650	754	1063	419	954
439	286	513	544	291	576
169	128	174	235	107	208
169	155	176	228	91	176
136	98	147	207	79	162

TWO-WAY TABLES

X(IJ**)				X(I*K*)			X(I**L)	
1084	63	120	559	361	691	774	885	941
784	134	200	1066	479	852	853	1034	1150
747	211	453	3053	1274	1817	1373	1797	2667
805	158	224	1462	725	1057	867	1243	1406
236	44	94	647	297	409	315	450	571
132	34	80	749	324	404	267	436	559
157	34	61	577	234	354	241	362	467

X(*JK*)			X(*J*L)		X(**KL)	
578	1410	1957	1918	2027	2054	1640
117	277	284	141	537	2544	3040
217	503	512	341	891	1609	3081
2782	3394	1937	3807	4306		

ONE-WAY TABLES

X(I***)
1826 2184 4464 2649 1021 995 829

X(*J**)
3945 678 1232 8113

X(**K*)
3694 5584 4690

X(***)L)
6207 7761

TOTAL

X(****)
13968

PRINT OF SUMS

SUM 2X(IJKL)LNX(IJKL)=.1419490011041468 + 006

SUM 2X(IJK*)LNX(IJK*)=.1602167221282754 + 006

SUM 2X(*JKL)LNX(*JKL)=.1900742436081673 + 006

SUM 2X(IJ*L)LNX(IJ*L)=.1706458462114764 + 006

SUM 2X(I*KL)LNX(I*KL)=.1680009679212620 + 006

SUM 2X(IJ**)LNX(IJ**)=.1893749331578995 + 006

SUM 2X(I*K*)LNX(I*K*)=.1866950003995916 + 006

SUM 2X(I**L)LNX(I**L)=.1976933167165831 + 006

SUM 2X(*JK*)LNX(*JK*)=.2085192406453136 + 006

SUM 2X(*J*L)LNX(*J*L)=.2189304551608272 + 006

SUM 2X(**KL)LNX(**KL)=.2175298383613909 + 006

SUM 2X(I****)LNX(I****)=.2168255594020366 + 006

SUM 2X(*J**)LNX(*J**)=.2377594145686747 + 006

SUM 2X(**K*)LNX(**K*)=.2363330848251296 + 006

SUM 2X(***)LNX(***)L)=.2474453182056570 + 006

2N LN N=.2666358302324256 + 006

TABLE OF NORMALIZED RESIDUALS

R(IJ11)				R(IJ12)			
-1.162	-.213	2.938	-.481	.155	.113	.552	.616
.671	-.308	1.295	.579	.277	.769	-1.494	-1.404
-1.619	2.262	-2.707	-.476	.664	.178	-.408	1.308
.484	-.156	-.868	.881	.254	-.385	.913	-1.403
1.779	-.916	1.954	-.627	-.734	-.323	-.278	-.060
.116	.062	-.404	-.499	.007	-1.722	1.610	.478
1.519	.996	-.029	.108	-2.067	-1.552	-.246	.216
R(IJ21)				R(IJ22)			
.155	-.947	-.548	-.152	-.434	2.540	-.713	.244
.523	.119	.212	-.251	-.195	.249	.112	-.300
-.835	-.963	-.796	1.344	1.829	-.673	-.580	-.803
-.201	-.207	1.106	.181	-.277	-.329	-.032	-.003
-.139	.523	-1.305	-.402	-1.039	-.831	1.893	.728
-.091	1.735	1.771	-.699	-1.064	.552	-.345	.408
1.925	1.042	1.940	-2.536	-.863	-.754	-.857	1.736
R(IJ31)				R(IJ32)			
1.220	-1.218	-.599	-.225	-.281	-1.242	-.326	.060
-.253	-.737	.162	-1.584	-.527	-.227	-.141	2.448
-.889	-.329	-.084	2.053	.416	.302	2.438	-2.597
-1.224	.555	.004	-.722	1.011	.485	-.801	.675
-.190	-.337	-.198	.842	.373	1.238	-1.661	-.247
.248	.134	-.808	1.027	.659	-.404	-1.182	-.355
.800	.092	-.319	-.359	-1.417	.724	-.063	.653

2Z LN Z = .1417767440958978 + 006

CHI-SQUARED = .1724668778 + 003

SECOND-ORDER INTERACTION = .1722570082 + 003

TABLE OF NORMALIZED RESIDUALS

R(IJ11)				R(IJ12)			
-.384	.383	1.070	-.079	.520	-.216	-.943	.094
-.524	-.937	.835	.290	.818	.763	-.852	-.387
.291	1.084	-.593	-.153	-.280	-.694	.352	.157
.018	-.598	-1.085	.363	-.024	.495	.952	-.463
.620	-.773	1.245	-.509	-.848	.573	-.978	.594
.143	.541	-.863	.097	-.170	-.495	.656	-.103
.513	.491	-.550	-.156	-.953	-.560	.579	.177
R(IJ21)				R(IJ22)			
.023	-.422	-.429	.214	-.026	.174	.332	-.210
-.278	.430	-.736	.417	.334	-.233	.598	-.418
-.150	-.600	.541	.023	.131	.299	-.300	-.021
.103	-.140	.297	-.141	-.113	.081	-.218	.142
.316	1.063	-1.518	.132	-.333	-.523	.947	-.115
.379	-.037	.937	-.421	-.392	.025	-.640	.378
-.038	.252	.879	-.324	.049	-.172	-.665	.249
R(IJ31)				R(IJ32)			
.208	.267	-.623	-.201	-.188	-.082	.363	.158
.628	.484	.210	-1.009	-.532	-.177	-.116	.732
-.037	-.462	-.179	.190	.026	.179	.077	-.142
-.121	.824	.716	-.369	.095	-.328	-.364	.275
-.777	-.452	.750	.608	.655	.171	-.362	-.439
-.451	-.339	-.398	.493	.367	.176	.207	-.358
-.329	-.703	-.656	.786	.329	.356	.370	-.479

2W LN W = .1419042077670569 + 006

CHI-SQUARED = .4418652549 + 002

THIRD-ORDER INTERACTION = .4479333708 + 002

7. References

- Anscombe, F. J. and Tukey, J. W. (1963), The examination and analysis of residuals, *Technometrics*, **5**, 141-160.
- Barnard, G. A. (1947), Significance tests for 2×2 tables, *Biometrika*, **34**, 139-167.
- Bartlett, M. S. (1935), Contingency table interactions, *J. Roy. Statist. Soc. Supplement* **2**, 248-252.
- Bhapkar, V. P. (1961), Some tests for categorical data, *Ann. Math. Statist.* **32**, 72-83.
- Bhapkar, V. P. and Koch, G. G. (1965), On the hypothesis of "no interaction" in three-dimensional contingency tables, Institute of Statistics Mimeo Series No. 440, University of North Carolina.
- Bhapkar, V. P. and Koch, G. G. (1966), Hypothesis of no interaction in four-dimensional contingency tables, Institute of Statistics Mimeo Series No. 449, University of North Carolina.
- Bhapkar, V. P. and Koch, G. G. (1968), Hypothesis of "no interaction" in multidimensional contingency tables, *Technometrics*, **10**, 107-123.
- Birch, M. W. (1963), Maximum likelihood in three-way contingency tables, *J. Roy. Statist. Soc. B*, **25**, 220-233.
- Brown, D. T. (1959), A note on approximations to discrete probability distributions, *Information and Control*, **2**, 386-392.
- Cox, D. R. and Lauh, Elizabeth (1967), A note on the graphical analysis of multidimensional contingency tables, *Technometrics* **9**, 481-488.
- Darroch, J. N. (1962), Interactions in multi-factor contingency tables, *J. Roy. Statist. Soc. B*, **24**, 251-263.
- Deming, W. E. and Stephan, F. F. (1940), On a least squares adjustment of a sampled frequency table when the expected marginal totals are known, *Ann. Math. Statist.* **11**, 427-444.
- Good, I. J. (1963), Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables, *Ann. Math. Statist.* **34**, 911-934.
- Good, I. J. (1965), The Estimation of Probabilities: An Essay on Modern Bayesian Methods, Research Monographs No. 30, The M. I. T. Press, Cambridge, Mass.
- Good, I. J. (1966), How to estimate probabilities, *J. Inst. Math. Applications* **2**, 364-383.
- Goodman, L. A. (1964a), Interactions in multidimensional contingency tables, *Ann. Math. Statist.*, **35**, 632-646.
- Goodman, L. A. (1964b), Simple method for analyzing three-factor interaction in contingency tables, *J. Amer. Statist. Assoc.* **59**, 319-352.
- Hoyt, C. J., Krishnaiah, P. R., and Torrance, E. P. (1959), Analysis of complex contingency data, *J. Exp. Ed.* **27**, 187-194.
- Ireland, C. T. and Kullback, S. (1968), Contingency tables with given marginals, *Biometrika*, **55**, 179-188.
- Kastenbaum, M. A. and Lamphiear, D. E. (1959), Calculation of chi-square to test the no three-factor interaction hypothesis, *Biometrics* **15**, 107-115.
- Kullback, S. (1959), *Information Theory and Statistics* (John Wiley & Sons, Inc., New York, N.Y.) (Dover Publications, Inc., New York, N.Y., 1968).
- Kullback, S. and Khairat, M. A. (1966), A note on minimum discrimination information, *Ann. Math. Statist.* **37**, 279-280.
- Kullback, S., Kupperman, M., and Ku, H. H. (1962a), Tests for contingency tables and Markov chains, *Technometrics* **4**, 573-608.
- Kullback, S., Kupperman, M., and Ku, H. H. (1962b), An application of information theory to the analysis of contingency tables with a table of $2N \ln N$, $N = 1(1) 10,000$, *J. Res. NBS* **66B** (Math. and Math. Phys.), No. 4, 217-243.
- Kupperman, M. (1957), Further applications of information theory to multivariate analysis and statistical inference, Dissertation, Graduate Council of George Washington University.
- Lancaster, H. O. (1951), Complex contingency tables treated by the partition of chi-square, *J. Roy. Statist. Soc. B* **13**, 242-249.
- Lewis, B. N. (1962), On the analysis of interaction in multidimensional contingency tables, *J. Roy. Statist. Soc. A*, **125**, 88-117.
- Lewis, P. M. II (1959), Approximating probability distributions to reduce storage requirements, *Information and Control*, **2**, 214-225.
- Lindley, D. V. (1964), The Bayesian analysis of contingency tables, *Ann. Math. Statist.* **35**, 1622-1643.
- Mantel, N. (1966), Models for complex contingency tables and polychotomous dosage response curves, *Biometrics* **22**, 83-95.
- Neyman, J. (1949), Contributions to the theory of the chi-square test, Berkeley Symposium on Mathematical Statistics and Probability, Ed. J. Neyman, Vol. **1**, 239-273, Univ. of Calif. Press, Berkeley.
- Norton, H. W. (1945), Calculation of chi-square for complex contingency tables, *J. Am. Statist. Assoc.* **40**, 251-258.
- Plackett, R. L. (1962), A note on interaction in contingency tables, *J. Roy. Statist. Soc. B* **24**, 162-166.
- Rao, C. R. (1965), *Linear Statistical Inference and its Applications* (John Wiley & Sons, Inc., New York, N.Y.).
- Ries, P. N. and Smith, H. (1963), The use of chi-square for preference testing in multidimensional problems, *Chemical Engineering Progress* **59**, 39-43.
- Roy, S. N. and Kastenbaum, M. A. (1956), On the hypothesis of no "interaction" in a multiway contingency table, *Ann. Math. Statist.* **27**, 749-757.
- Simpson, C. H. (1951), The interpretation of interaction in contingency tables, *J. Roy. Statist. Soc. B* **13**, 238-241.
- Wald, A. (1943), Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Trans. Am. Math. Soc.* **54**, 426-482.
- Woolf, B. (1955), On estimating the relation between blood group and disease, *Annals of Human Genetics* **19**, 251-253.

(Paper 72B3-269)